Introduction
class="introduction"
The module introduces the discipline of economics.
Do You Use Facebook?

Economics is greatly impacted by how well information travels through society. Today, social media giants Twitter, Facebook, and Instagram are major forces on the information super highway. (Credit: Johan Larsson/Flickr)

Note:

Decisions ... Decisions in the Social Media Age

Economics is about choices and decisions

To post or not to post a comment? Should I change my status? Every day we are faced with many decisions like these, from the very basic ones like what to wear, what to have for breakfast, which route to take to university, to the more complex choices like what career to follow and what course to study. Our responses to these choices depends on the information we have available at any given moment; information economists call “imperfect” because we rarely have all the data we need to make perfect decisions. Despite the lack of perfect information, we still make hundreds of decisions a day.

And now we have another means with which to gather information—the internet and social media in particular. Apps like Facebook, WhatsApp, Twitter and even Tinder (online dating) are altering the process by which we make choices, how we spend our time, which movies we see, which

products we buy, who we interact with and more. How many of you chose a university without first checking out its Website, Facebook page or Twitter stream first for information and feedback?

As you will see in this course, what happens in economics (and, indeed, in life) is affected by how well and how fast information is disseminated through a society. The internet (including social media applications and search engines) allows information to be accessed almost instantly by anyone at virtually no cost. Does this fact make it easier or more difficult to make choices and decisions? Certainly the world's problems do not seem to be any less serious in spite of the enormous amounts of information we seem to have available to us these days. Globally we still have mass poverty, economic inequality, unemployment, social disruption, wars, growing environmental destruction and global warming. Are these problems the result of bad choices and poor decisions? In economics (as in life) good choices and decisions must surely be based on all available information.

This leads us to the topic of this chapter, an introduction to the world of making decisions, processing information, and understanding behaviour in markets —the world of economics.

Note:

Introduction

In this chapter, you will learn about:

- What Economics Is About And Why It Is Important
- Microeconomics and Macroeconomics
- How Economists Use Theories and Models to Understand Economic Issues
- How Economies Can Be Organized: An Overview of Economic Systems

What is economics and why should you spend your time studying it? After all, there are other disciplines you could be studying, and other ways you

could be spending your time. Well, as we have seen from the introduction, economics is actually about life. What is life if it is not a series of choices that take people on their different paths from the cradle to the grave? If you are someone who makes choices, in whatever sphere of life, then studying economics will be just about the best choice/decision you will ever make.

Economics: The science of choice

Economics is probably not what you think. It is not primarily about money or finance. It is not primarily about business. It is not mathematics. What is it then? It is both a discipline and a way of viewing the world. Economics encompasses not only knowledge. It offers a useful skills set to those intent on making better choices and decisions.

What is economics and why is it important?

By the end of this section, you will be able to:

- Discuss the importance of studying economics
- Explain the relationship between production and division of labor
- Evaluate the significance of scarcity

Economics is the study of how people make decisions in the face of scarcity. These can be individual decisions, family decisions, business decisions or societal decisions. Look around carefully and you will see that scarcity is a fact of life. **Scarcity** means that human wants for goods, services and resources exceed what is available. Resources, such as labor, tools, land, and raw materials are necessary to produce the goods and services we want but their supply is limited. Of course, the ultimate scarce resource is time- everyone, rich or poor, has just 24 hours in the day to try to acquire the goods they want. At any point in time, there is only a finite amount of resources available.

Think about it this way: According to its 2011 census (Statistics South Africa: 2012) South Africa has a population of nearly 52 million people. The total area of South Africa is 1 219 090 square kilometers of which only 12.1% is arable (suitable for growing crops). Even though South Africa is relatively rich in natural resources (especially mineral resources) it is a drought prone country. The threat of future water shortages poses a real problem for the country. Because these resources (not only natural resources but also labor, capital and entrepreneurship) are limited, so are the numbers of goods and services we can produce with them. Combine this with the fact that human wants seem to be virtually infinite, and you can see why scarcity is a problem.

Scarcity of Resources

Homeless people are a stark reminder that scarcity of resources is real. (Credit: “daveynin”/Flickr Creative Commons)

If you still do not believe that scarcity is a problem, consider this: Does everyone need food to eat? Does everyone need a decent place to live? Does everyone have access to healthcare? In every country in the world, there are people who are hungry, homeless (for example, those who call park benches their beds, as shown in [link]), and in need of healthcare, who struggle to scratch out a living just to secure a few critical goods and services. Why is this the case? It is because of scarcity. Let’s look a little more closely at the concept of scarcity because it is crucial to understanding economics.

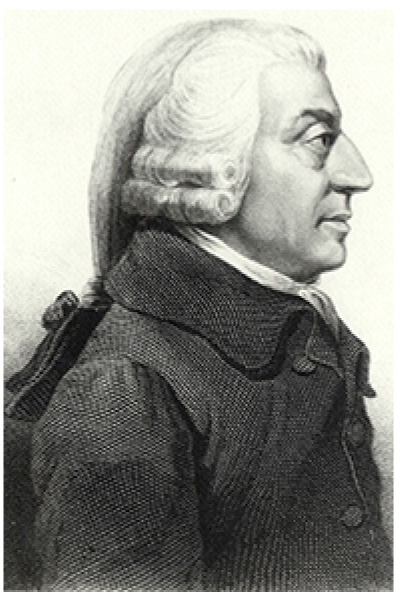
The Problem of Scarcity

Think about all the things you consume: food, shelter, clothing, transportation, healthcare, and entertainment. How do you acquire those items? You do not produce them yourself. You buy them. How do you afford the things you buy? You work for pay. Or if you do not, someone

else does on your behalf. Yet most of us never have enough to buy all the things we want. This is because of scarcity. So how do we solve it?

Every society, at every level, must make choices about how to use its resources. Families must decide whether to spend their money on a new car or a fancy vacation. Municipalities must choose whether to put more of the budget into maintenance and service delivery or to spend more on salaries. Nations must decide whether to devote more funds to national defense or to protecting the environment. In most cases, there just isn’t enough money in the budget to do everything. So why do we not each just produce all of the things we consume? The simple answer is most of us do not know how, but that is not the main reason. (When you study economics, you will discover that the obvious choice is not always the right answer—or at least the complete answer. Studying economics teaches you to think in a different of way.) Think back to earlier days, when people knew how to do so much more than we do today, from building their homes, to growing their crops, to hunting for food, to repairing their equipment. Most of us do not know how to do all—or any—of those things. It is not because we could not learn. Rather, we do not have to. The reason why is something called *the division and specialization of labor*, a production innovation first identified by Adam Smith, [link], in his book, *The Wealth of Nations*.

Adam Smith



Adam Smith

introduced the idea of dividing labor into discrete tasks. (Credit: Wikimedia Commons)

The Division of and Specialization of Labor

The formal study of economics began when Adam Smith (1723–1790) published his famous book (which is commonly called *The Wealth of Nations*) in 1776. Many authors had written on economics in the centuries before Smith, but he was the first to address the subject in a comprehensive way. In the first chapter, Smith introduces the **division of labor**, which means that the way a good or service is produced is divided into a number of tasks that are performed by different workers, instead of all the tasks being done by the same person.

To illustrate the division of labor, Smith counted how many tasks went into making a pin: drawing out a piece of wire, cutting it to the right length, straightening it, putting a head on one end and a point on the other, and packaging pins for sale, to name just a few. Smith counted 18 distinct tasks that were often done by different people—all for a pin, believe it or not!

Modern businesses divide tasks as well. Even a relatively simple business like a restaurant divides up the task of serving meals into a range of jobs like top chef, assistant chefs, less-skilled kitchen help, servers to wait on the tables, a greeter at the door, cleaners, and a business manager to handle payments and bills—not to mention the economic connections a restaurant has with suppliers of food, furniture, kitchen equipment, and the building where it is located. A complex business like a large manufacturing factory, such as the shoe factory shown in [link], or a hospital, can have hundreds of job classifications.

Division of Labor

Workers on an assembly line are an example of the divisions of labor. (Credit: Nina Hale/Flickr Creative Commons)

Why the Division of Labor Increases Production

When the tasks involved with producing a good or service are divided and subdivided, workers and businesses can produce a greater quantity of output. In his observations of pin factories, Smith observed that one worker alone might make 20 pins in a day, but that a small business of 10 workers (some of whom would need to do two or three of the 18 tasks involved with pin-making), could make 48,000 pins in a day. How can a group of workers, each specializing in certain tasks, produce so much more than the same number of workers who try to produce the entire good or service by themselves? Smith offered three reasons.

First, **specialization** in a particular small job allows workers to focus on the parts of the production process where they have an advantage. (In later chapters, we will develop this idea by discussing comparative advantage.) People have different skills, talents, and interests, so they will be better at some jobs than at others. The particular advantages may be based on educational choices, which are in turn shaped by interests and talents. Only

those with medical degrees qualify to become doctors, for instance. For some goods, specialization will be affected by geography—it is more realistic to be a wheat farmer in the Free State than on the North Coast of KwaZulu-Natal, but perhaps more suitable to run a tourist resort on the North Coast of KwaZulu-Natal (e.g. Ballito) than in the farming area of the Free State. If you live in or near a big city, it is easier to attract enough customers to operate a successful dry cleaning business or movie theater than if you live in a sparsely populated rural area. Whatever the reason, if people specialize in the production of what they do best, they will be more productive than if they produce a combination of things, some of which they are good at and some of which they are not.

Second, workers who specialize in certain tasks often learn to produce more quickly and with higher quality. This pattern holds true for many workers, including assembly line laborers who build cars, stylists who cut hair, and doctors who perform heart surgery. In fact, specialized workers often know their jobs well enough to suggest innovative ways to do their work faster and better.

A similar pattern often operates within businesses. In many cases, a business that focuses on one or a few products (sometimes called its "core competency") is more successful than firms that try to make a wide range of products.

Third, specialization allows businesses to take advantage of **economies of scale**, which means that for many goods, as the level of production increases, the average cost of producing each individual unit declines. For example, if a factory produces only 100 cars per year, each car will be quite expensive to make on average. However, if a factory produces 50,000 cars each year, then it can set up an assembly line with huge machines and workers performing specialized tasks, and the average cost of production per car will be lower. The ultimate result of workers who can focus on their preferences and talents, learn to do their specialized jobs better, and work in larger organizations is that society as a whole can produce and consume far more than if each person tried to produce all of their own goods and services. The division and specialization of labor has been a force against the problem of scarcity.

Trade and Markets

Specialization only makes sense, though, if workers can use the pay they receive for doing their jobs to purchase the other goods and services that they need. In short, specialization requires trade.

You do not have to know anything about electronics or sound systems to play music—you just buy an iPod or MP3 player, download the music and listen. You do not have to know anything about artificial fibers or the construction of sewing machines if you need a jacket—you just buy the jacket and wear it. You do not need to know anything about internal combustion engines to operate a car—you just get in and drive. Instead of trying to acquire all the knowledge and skills involved in producing all of the goods and services that you wish to consume, the market allows you to learn a specialized set of skills and then use the pay you receive to buy the goods and services you need or want. This is how our modern society has evolved into a strong economy.

Trade, and the markets in which trading occurs, is one particular type of economic system. Economic systems are devised by people to answer the three basic economic questions that spring from the essential economic problem of scarcity of resources. These are: what to produce, how to produce and for who to produce. Societies have used various types of economic systems to answer these questions over the ages although the market system is the most dominant of these today.

Why Study Economics?

Now that we have a better idea of what economics is concerned with, let's quickly discuss why you are right to study it. Economics is not primarily a collection of facts to be memorized, though there are plenty of important concepts to be learned. Instead, economics is better thought of as a collection of questions to be answered or puzzles to be worked out. Most important, economics provides the tools and skills to work out those puzzles.

- Virtually every major problem facing the world today, from global warming, to world poverty, to the conflicts in Syria, Afghanistan, and Somalia, has an economic dimension. If you are going to be part of solving those problems, you need to be able to understand them. Economics is crucial.
- Understanding economics is important for good citizenship and holding government accountable. You need to be able to vote intelligently on budgets, regulations, and laws in general. For instance, when President Zuma replaced Mr Nene as finance minister in 2015 how did this affect the economy? Do you know?
- A basic understanding of economics makes you a well-rounded thinker. When you read articles about economic issues, you will understand and be able to evaluate the writer's argument. When you hear classmates, co-workers, or political candidates talking about economics, you will be able to distinguish between common sense and nonsense. You will find new ways of thinking about current events and about personal and business decisions, as well as current events and politics.

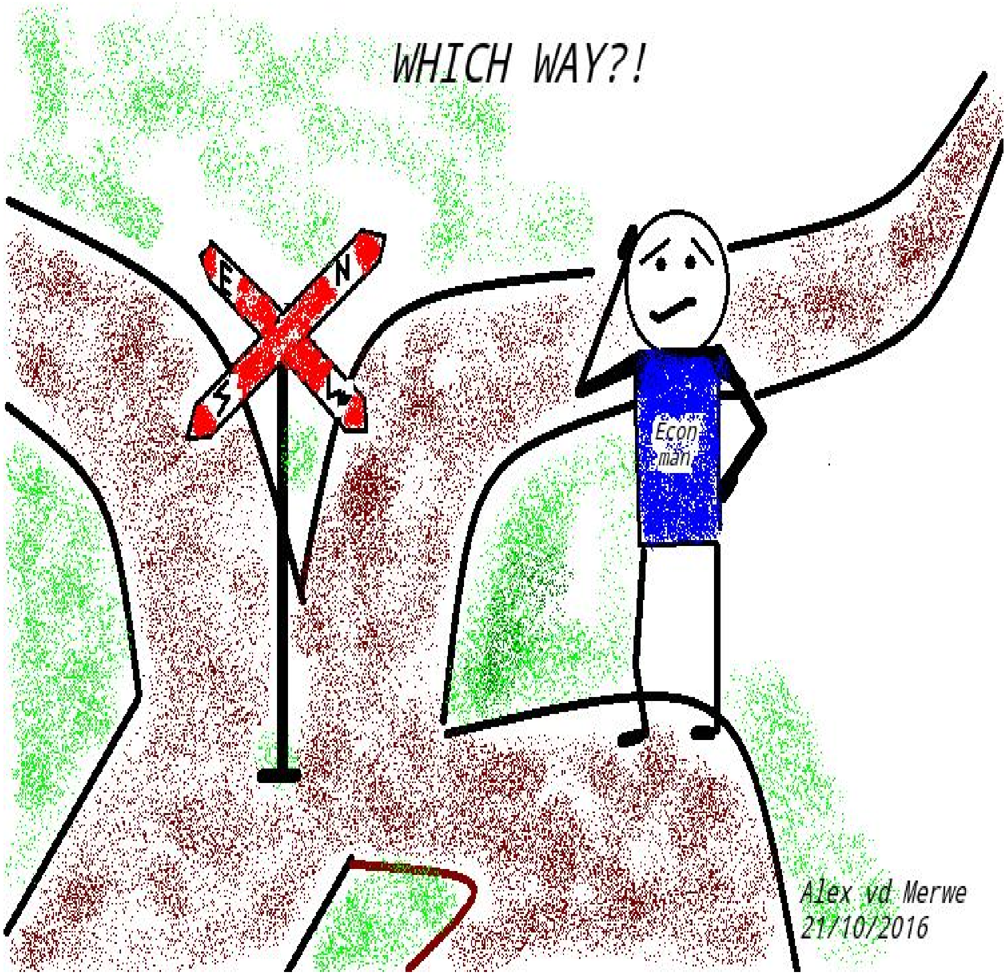


Figure 4: Economics is really just about the choices people make (Credit: Alex vd Merwe)

The study of economics does not dictate the answers, but it can illuminate the different choices.

Key Concepts and Summary

Economics seeks to solve the problem of scarcity, which is when human wants for goods and services exceed the available supply. A modern market economy displays a division of labor, in which people earn income by specializing in what they produce and then use that income to purchase the products they need or want. The division of labor allows individuals and

firms to specialize and to produce more for several reasons: a) It allows people and their businesses to focus on areas of advantage due to natural factors and skill levels; b) It encourages people to learn and invent; c) It allows individuals and firms to take advantage of economies of scale. Division and specialization of labor only work when individuals can purchase what they do not produce in markets. Learning about economics helps you understand the major problems facing the world today, prepares you to be a good citizen, and helps you become a well-rounded thinker.

Self-Check Questions

Exercise:

Problem: What is scarcity? Can you think of two causes of scarcity?

Solution:

Scarcity means human wants for goods and services exceed the available supply. Supply is limited because resources are limited. Demand, however, is virtually unlimited. Whatever the supply, it seems human nature to want more.

Exercise:

Problem:

Residents of the town of Nkandlaville like to consume 25 litre barrels of special putu beer, but each barrel requires one month and 10 people to produce it. If the town has a total of 100 people, what is the maximum amount of barrels the residents can consume in a month?

Solution:

100 people / 10 people per barrel = a maximum of 10 barrels per month if all residents produce barrels of putu beer. Since consumption is limited by production, the maximum number of barrels residents could consume per month is 10.

Exercise:

Problem:

A consultant works for R200 per hour. She likes to eat vegetables, but is not very good at growing them. Why does it make more economic sense for her to spend her time at the consulting job and shop for her vegetables?

Solution:

She is very productive at her consulting job, but not very productive growing vegetables. Time spent consulting would produce far more income than what she could save growing her vegetables using the same amount of time. So on purely economic grounds, it makes more sense for her to maximize her income by applying her labor to consulting which is what she does best (i.e. specialization of labor).

Exercise:

Problem:

A computer systems engineer could paint his house, but it makes more sense for him to hire a painter to do it. Explain why.

Solution:

The engineer is better at computer science than at painting. Thus, his time is better spent working for pay at his job and paying a painter to paint his house. Of course, this assumes he does not paint his house for fun!

Review Questions

Exercise:

Problem:

Give the three reasons that explain why the division of labor increases an economy's level of production.

Exercise:

Problem: What are three reasons to study economics?

Critical Thinking Questions

Exercise:

Problem:

Suppose you have a team of two workers: one is a baker and one is a butcher. Explain why the kitchen can produce more meals in a given period of time if each worker specializes in what they do best than if each worker tries to do everything from appetizer to dessert.

Exercise:

Problem: Why would division of labor without trade not work?

Exercise:

Problem:

Can you think of any examples of *free* goods, that is, goods or services that are not scarce?

Exercise 10
List four categories of factors of production or resources and give three examples of each.

References

Smith, A. 1904. An inquiry into the nature and causes of the wealth of nations. 5th ed. London: Methuen. Available: http://www.econlib.org/library/Smith/smWN.html?PHPSESSID=5eabcf2225ef1e63e8942c117d569496 (Accessed: 17 Oct 2016)

Statistics South Africa. 2012. Census 2011. Census in brief. Pretoria: Statistics South Africa. Report no. 03-01-41. Available: http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf (Accessed: 17 Oct 2016)

Glossary

division of labor
: the way in which the work required to produce a good or service is divided into tasks performed by different workers

economics
: the study of how humans make choices under conditions of scarcity

economies of scale
: when the average cost of producing each individual unit declines as total output increases

scarcity
: when human wants for goods and services exceed the available supply

specialization
: when workers or firms focus on particular tasks for which they are well-suited within the overall production process

Microeconomics and Macroeconomics

By the end of this section, you will be able to:

- Describe microeconomics
- Describe macroeconomics
- Contrast monetary policy and fiscal policy

Economics is concerned with the well-being of *all* people, including those with jobs and those without jobs, as well as those with high incomes and those with low incomes. Economics acknowledges that production of useful goods and services can create problems of environmental destruction. It explores the question of how investing in education helps to develop workers' skills. It asks questions like how to tell when big businesses or big labor unions are operating in a way that benefits society as a whole and when they are operating in a way that benefits their owners or members at the expense of others. It looks at how government spending, taxes, and regulations affect decisions about production and consumption.

It should be clear by now that economics covers a lot of ground. That ground can be divided into two parts: **Microeconomics** focuses on the actions of individual agents within the economy, like households, workers, and businesses; **Macroeconomics** looks at the economy as a whole. It focuses on broad issues such as growth of production, the number of unemployed people, the inflationary increase in prices, government deficits, and levels of exports and imports. Microeconomics and macroeconomics are not separate subjects. They just offer different economic points of view: one narrower, the other broader.

To understand why both microeconomic and macroeconomic perspectives are useful, consider the problem of studying a biological ecosystem like a lake. One person who sets out to study the lake might focus on specific topics: certain kinds of algae or plant life; the characteristics of particular fish or snails; or the trees surrounding the lake. Another person might take an overall view and instead consider the entire ecosystem of the lake from top to bottom; what eats what, how the system stays in a rough balance, and what environmental stresses affect this balance. Both approaches are useful, and both examine the same lake, but the viewpoints are different. In a

similar way, both microeconomics and macroeconomics study the same economy, but each has a different viewpoint.

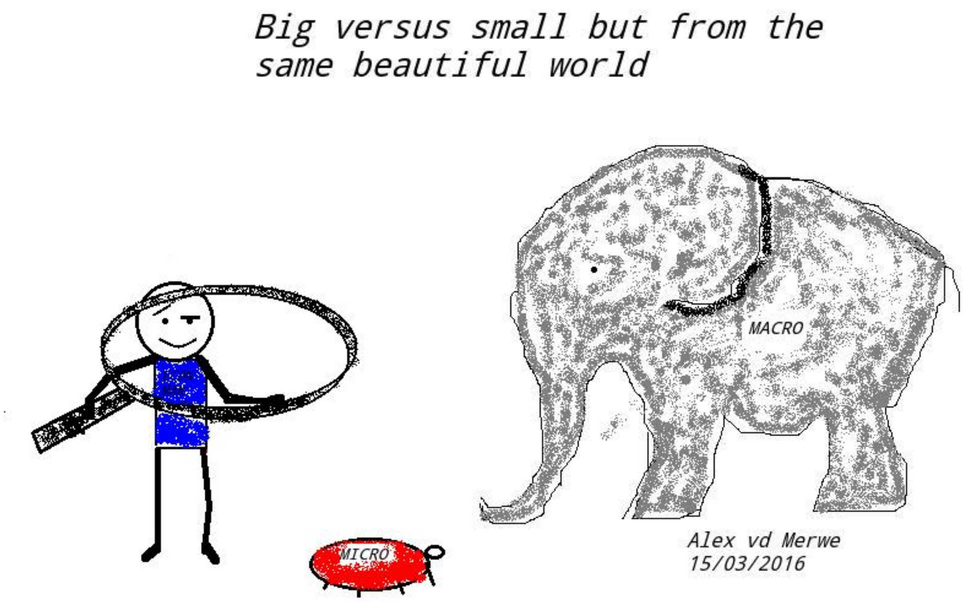


Figure 1

Whether you are looking at lakes or economics, the micro and the macro insights should blend with each other. In studying a lake, the micro insights about particular plants and animals help to understand the overall food chain, while the macro insights about the overall food chain help to explain the environment in which individual plants and animals live.

In economics, the micro decisions of individual businesses are influenced by whether the macroeconomy is healthy; for example, firms will be more likely to hire workers if the overall economy is growing. In turn, the performance of the macroeconomy ultimately depends on the microeconomic decisions made by individual households and businesses.

Microeconomics

Microeconomics, from the perspective of consumers for example, is typically concerned with questions such as what determines how households and individuals spend their budgets? What combination of goods and services will best fit their needs and wants, given the budget they have to spend? How do people decide whether to work, and if so, whether to work full time or part time? How do people decide how much to save for the future, or whether they should borrow to spend beyond their current means?

From the perspective of producers, microeconomics would be concerned with issues such as, for example, what determines the products, and how many of each, a firm will produce and sell? What determines what prices a firm will charge? What determines how a firm will produce its products? What determines how many workers it will hire? How will a firm finance its business? When will a firm decide to expand, downsize, or even close? In the microeconomics part of this book, we will learn about the theory of consumer behavior and the theory of the firm.

Macroeconomics

Macroeconomics is concerned with questions such as, for example, what determines the level of economic activity in a society? In other words, what determines how many goods and services a nation actually produces? What determines how many jobs are available in an economy? What determines a nation's standard of living? What causes the economy to speed up or slow down? What causes firms to hire more workers or to lay workers off? Finally, what causes the economy to grow over the long term?

An economy's macroeconomic health can be defined by a number of goals: growth in the standard of living (which requires economic growth), low unemployment, and low inflation, to name the most important. How can macroeconomic policy be used to pursue these goals? **Monetary policy**, which involves policies that affect bank lending, interest rates, and financial capital markets, is conducted by a nation's central bank. For South Africa, this is the South African Reserve Bank in Pretoria. **Fiscal policy**, which involves government spending and taxes, is directed by the Minister of Finance together with the National Treasury. South Africa's fiscal policy is

reflected in the National Budget which is announced at the end of February each year. These are the main tools the government has to work with. South Africans tend to expect that government can fix whatever economic problems we encounter, but to what extent is that expectation realistic? These are just some of the issues that will be explored in the macroeconomic chapters of this book.

Key Concepts and Summary

Microeconomics and macroeconomics are two different perspectives on the economy. The microeconomic perspective focuses on parts of the economy: individuals, firms, and industries. The macroeconomic perspective looks at the economy as a whole, focusing on goals like growth in the standard of living, unemployment, and inflation. Macroeconomics has two types of policies for pursuing these goals: monetary policy and fiscal policy.

Self-Check Questions

Exercise:

Problem:

What would be another example of a “system” in the real world that, like macro and microeconomics, distinguishes between smaller and bigger parts of the same system?

Solution:

There are many physical systems that would work, for example, the study of planets (micro) in the solar system (macro), or solar systems (micro) in the galaxy (macro).

Review Questions

Exercise:

Problem:

What is the difference between microeconomics and macroeconomics?

Exercise:

Problem: What are examples of individual economic agents?

Exercise:

Problem: What are the three main goals of macroeconomics?

Critical Thinking Questions

Exercise:

Problem:

A balanced national budget (tax revenues equal to government expenditure) and a balance of payments (import payment and investment outflows balanced by export receipts and investment inflows) are considered secondary goals of macroeconomics, while growth in the standard of living (for example) is considered a primary goal. Why do you think that is so?

Exercise:

Problem:

Macroeconomics is an aggregate (total) of what happens at the microeconomic level. Would it be possible for what happens at the macro level to differ from how economic agents would react to some stimulus at the micro level? *Hint:* Think about the behavior of crowds.

Glossary

fiscal policy
: economic policies that involve government spending and taxes

macroeconomics
: the branch of economics that focuses on broad issues such as growth, unemployment, inflation, and trade balance.

microeconomics
: the branch of economics that focuses on actions of particular agents within the economy, like households, workers, and business firms

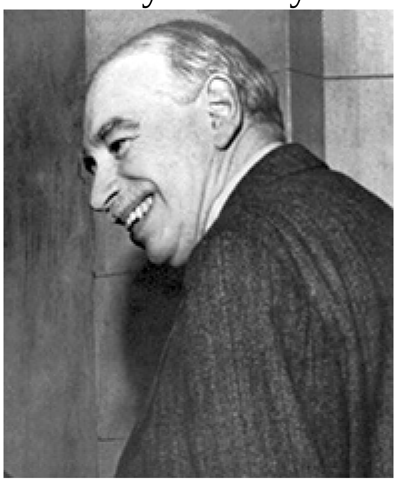
monetary policy
: policy that involves altering the level of interest rates, the availability of credit in the economy, and the extent of borrowing

How Economists Use Theories and Models to Understand Economic Issues

By the end of this section, you will be able to:

- Interpret a circular flow diagram
- Explain the importance of economic theories and models
- Describe goods and services markets and labor markets

John Maynard Keynes



One of the most influential economists in modern times was John Maynard Keynes. (Credit: Wikimedia Commons)

John Maynard Keynes (1883–1946), one of the greatest economists of the twentieth century, pointed out that economics is not just a subject area but also a way of thinking. Keynes, shown in Figure 1, famously wrote in the introduction to the Cambridge Economics Handbooks series: “[Economics] is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions.” (Keynes:

1922: v.) In other words, economics teaches you how to think, not what to think.

Note:
Watch this video about John Maynard Keynes and his influence on economics.

Economists view the world differently to practitioners of other disciplines. They analyze issues and problems using economic theories that are based on particular **assumptions** about human behavior, that are different from the assumptions an anthropologist or psychologist might use. A **theory** is a simplified representation of how two or more variables interact with each other. The purpose of a theory is to take a complex, real-world issue and simplify it down to its essentials. If done well, this enables the analyst to understand the issue and any problems around it. A good theory is simple enough to be understood, while complex enough to capture the key features of the object or situation being studied.

Sometimes economists use the term **model** instead of theory. Strictly speaking, a theory is a more abstract representation, while a model is more applied or empirical representation. Models are used to test theories, but for this course we will use the terms interchangeably.

Figure 2: A model of a fighter jet (Credit: FlickR Creative Commons by 2.0)

For example, an architect who is planning a major office building will often build a small-scale physical model to show how the entire city block will look after the new building is constructed. Companies often build models of their new products, which are more rough and unfinished than the final product will be, but can still demonstrate how the new product will work. Motor and aircraft manufacturers also usually first design and build small-scale models to get some idea of the end product. In the same way economics relies on simple models to explain more complex ideas.

Economics is a social science

Economics is a social science since it is concerned with the choices people (society) make. Economics examines the behaviour of individuals (consumers) and institutions (firms and government) engaged in economic activity, that is, production, trade, consumption and the generation of income. Economists seek principles/generalisations about the way individuals and institutions behave. Economic principles and theories are sensible statements about economic behaviour or the economy.

Economics has its own special words (terminology) which you should become familiar with. This is why a special glossary is provided at the end of most sections and chapters in this book. In many instances there may be different words for the same concept. For example, laws, principles, theories and models all mean the same thing. They are generalizations about the economic behavior of individuals and institutions. Since there are billions of people and they are all unique, it is firstly impossible to study

each person's behavior and, secondly, it is equally impossible to understand or predict each one's choices/decisions with a great degree of accuracy. We can only get a general idea or understanding of an individuals' behavior. That is the best we can do. It is in this sense that economics is different from the natural sciences such as physics, chemistry and biology. In these disciplines it is possible to achieve at least close to 100% accuracy and predictability in experiments. And there is generally only one correct answer to scientific questions. This is not the case in economics since people are all different and so their behavior is not 100% predictable.

Assumptions help economists to generalize.The most important assumption in economics that allows economists to draw sensible conclusions about peoples' behavior is that of "ceteris paribus". This is a Latin phrase meaning "all else being equal" or "other things held constant". So, for example, we could predict that Sifiso will have more income to spend if he gets a promotion, ceteris paribus (that is, assuming that nothing unusual happens that will prevent this such as, for example, getting ill and having to spend more on health care). The ceteris paribus assumption allows us to focus just on the issues we are interested in without complicating things by ignoring all else that could affect the relationship.

Since economics is a social science, allowance has to be made for the fact that people may perceive things differently. So people, because they have their own beliefs and attitudes, often make different choices and decisions even if they are faced with the same situation. For example, a salary increase of R1000 per month might be regarded as very satisfactory for a low income earner, but a very rich person would consider the same R1000 as very unsatisfactory. And so economics can be divided into **normative economics** (to do with opinions and subjective beliefs) and **positive economics**. Positive economics has to do with economic facts. In our example of the R1000 salary increase the fact is that all workers (rich and poor) got the same R1000 increase. That poor workers regarded the increase as satisfactory but the rich though it not good enough is a normative economic issue.

The laws, principles, theories and models that are produced by economists with the help of generalizations such as "ceteris paribus" may be expressed

in different ways: Written descriptions, tables or schedules, graphs and pictures (images). Economists are particularly fond of graphs. However, images or simple pictures such as the circular flow model of economic activity also tell a story.

A good model to start with in economics is the **circular flow diagram**, which is shown in Figure 3. It pictures the economy as consisting of two groups—households and firms—that interact in two markets: the **goods and services market** in which firms sell and households buy and the **labor market** in which households sell labor, capital and land (resources) to business firms or other employees.

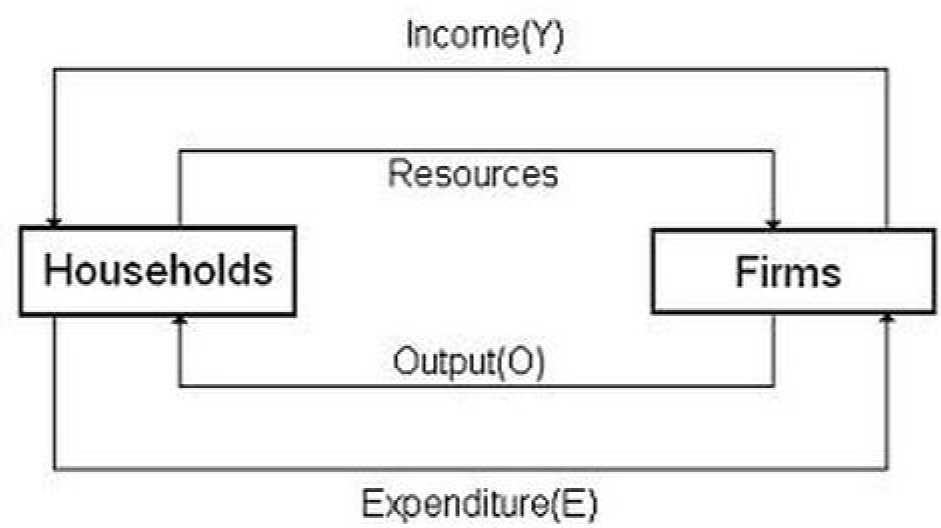


Figure 3: The circular flow diagram shows how households and firms interact in the goods and services market, and in the factor (resources) market. The direction of the arrows shows that in the goods and services market, households receive goods and services and pay firms for them. In the factor (resources) market, households provide

labor and receive payment from firms through wages, salaries, and benefits.They also receive rent for their land, and interest payments for capital.This is the simple 2-sector (households and firms) circular flow model.

Of course, in the real world, there are many different markets for goods and services and markets for the different types of factors of production. The circular flow diagram simplifies this to make the picture easier to grasp. In the diagram, firms produce goods and services, which they sell to households in the product market in return for revenues. Households sell their labor, land,capital and entrepreneurial skill to firms in the factor market in return for wages/salaries, rent, interest and profit respectively. This particular model does not even show the product and factor markets. Go ahead and draw them in to show a 2-sector, 2-market market economy.

This version of the circular flow model is stripped down to the essentials, but it has enough features to explain how the the two main sectors of any market economy interact through the product and factor markets. We could easily add details to this basic model if we wanted to introduce more real-world elements, like financial markets, governments, and interactions with the rest of the globe (imports and exports).

Economists have their own methods or tools to carry out their investigations. When they see an economic issue or problem, they go through the theories they know to see if they can find one that fits. Then they use the theory to see if it fits the issue or problem. In economics, theories may be expressed as diagrams, graphs, or even as mathematical equations. (Do not worry. In this course, we will mostly use only graphs.) Economists do not figure out the answer to the problem first and then draw the graph to illustrate. Rather, they use the graph of the theory to help them figure out the answer. Although at the introductory level, you can sometimes figure out the right answer without applying a model, if you keep studying economics, before too long you will run into issues and problems that you will need to graph to solve. Both micro and macroeconomics are explained in terms of theories and models. The most

well-known theories are probably those of supply and demand, but you will learn a number of others.

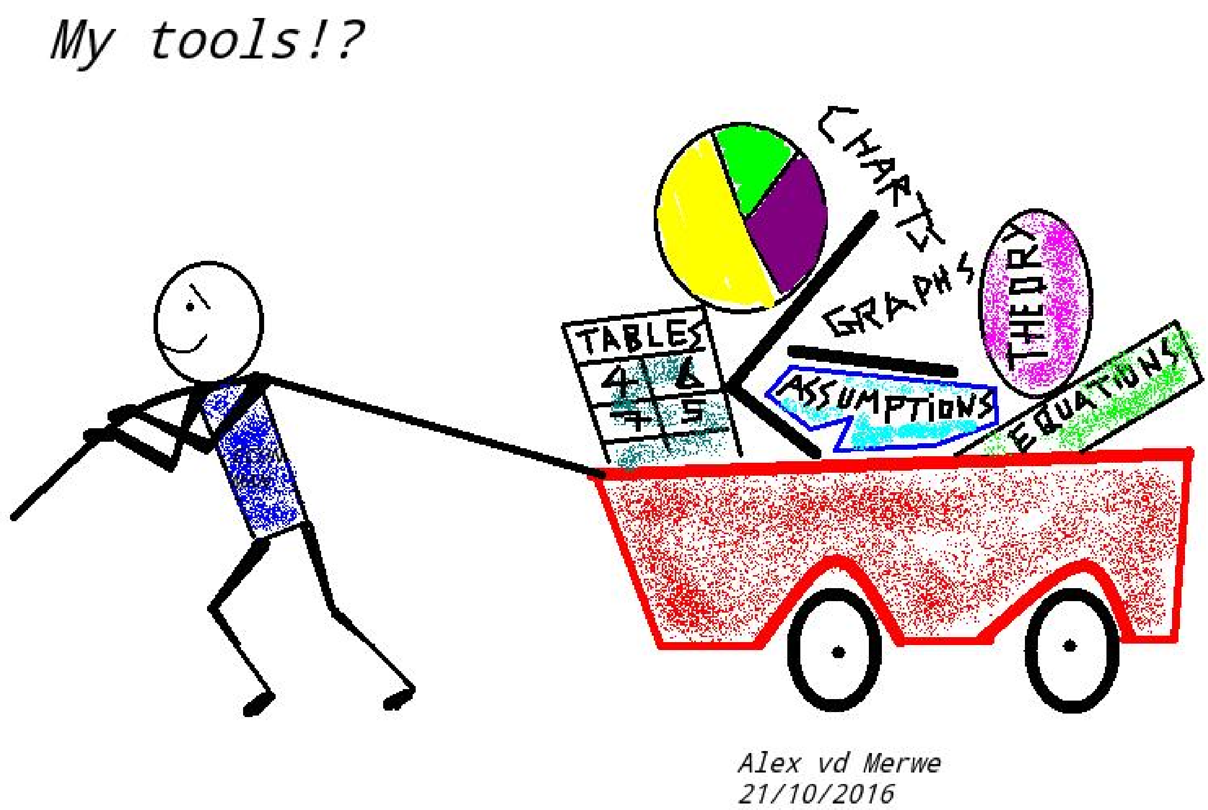


Figure 4: Some economics tools.

Key Concepts and Summary

Economists analyze problems differently than do other disciplinary experts. The main tools economists use are economic theories or models. A theory is not an illustration of the answer to a problem. Rather, a theory is a tool for determining the answer.

Self-Check Questions

Exercise:

Problem:

Suppose we extend the simple 2-sector circular flow model (complete with product and factor markets) to add imports and exports. Copy the circular flow diagram onto a sheet of paper and then add a foreign country as a third agent. Draw a rough sketch of the flows of imports, exports, and the payments for each on your diagram.

Solution:

Draw a box outside the original circular flow to represent the foreign country. Draw an arrow from the foreign country to firms, to represents imports. Draw an arrow in the reverse direction representing payments for imports. Draw an arrow from firms to the foreign country to represent exports. Draw an arrow in the reverse direction to represent payments for imports.

Exercise:

Problem:

What is an example of a problem in the world today, not mentioned in the chapter, that has an economic dimension?

Solution:

There are many such problems. Consider the AIDS epidemic. Why are so few AIDS patients in Africa and Southeast Asia treated with the same drugs that are effective in the United States and Europe? It is because neither those patients nor the countries in which they live have the resources to purchase the same drugs.

Review Questions

Exercise:

Problem: How did John Maynard Keynes define economics?

Exercise:

Problem:

Are households primarily buyers or sellers in the goods and services market? In the labor market?

Exercise:

Problem:

Are firms primarily buyers or sellers in the goods and services market? In the labor market?

References

Keynes, J.M. 1922. Introduction to the Cambridge Economics Handbooks series. In: Robertson, D.H. 1922. Money. (Cambridge economics handbooks series). Cambridge: Cambridge University Press. page v.

Glossary

circular flow diagram
: a diagram that views the economy as consisting of households and firms interacting in a goods and services market and a labor market

goods and services market
: a market in which firms are sellers of what they produce and households are buyers

labor market
: the market in which households sell their labor as workers to business firms or other employers

model
: see theory

theory

a representation of an object or situation that is simplified while including enough of the key features to help us understand the object or situation

How Economies Can Be Organized: An Overview of Economic Systems

By the end of this section, you will be able to:

- Contrast traditional economies, command economies, and market economies
- Explain gross domestic product (GDP)
- Assess the importance and effects of globalization

Think about what a complex system a modern economy is. It includes all production of goods and services, all buying and selling, all employment. The economic life of every individual is interrelated, at least to a small extent, with the economic lives of thousands or even millions of other individuals. Who organizes and coordinates this system? Who insures that, for example, the number of televisions a society provides is the same as the amount it needs and wants? Who ensures that the right number of employees work in the electronics industry? Who ensures that televisions are produced in the best way possible - efficiently, safely, to a high standard? How does it all get done?

Figure 1: Like puzzles, economic systems are arranged in different ways.

A traditional economic system

Figure 2: Subsistence farming is an example of a traditional economic system.(Credit: Palmer, 2010)

There are at least three ways societies have found to organize an economy. The first is the **traditional economy**, which is the oldest economic system and can be found in parts of Asia, Africa, and South America. Traditional economies organize their economic affairs the way they have always done (i.e., tradition). Occupations stay in the family. Most families are farmers who grow the crops they have always grown using traditional methods for subsistence purposes. What you produce is what you get to consume. Because things are driven by tradition, there is little economic progress or development.

Figure 3: Ancient Egypt was an example of a command economy. (Credit: Jay Bergesen/Flickr Creative Commons)

Command economies are very different to other systems of economic organization. In a **command economy**, economic effort is focused on goals passed down from a ruler or ruling class. Ancient Egypt was a good example: a large part of economic life was focused on building pyramids for the pharaohs, like those shown in Figure 3. Medieval Europe is another example: the lord of the manor (farm) provided the land for growing crops and protection in the event of war. In return, vassals (slaves) provided labor and others served as soldiers in the lord's armies. In the last century, communism emphasized command economies. Communist countries that employed command economic systems in the last century included many Eastern European countries (including Russia) and China.

In a command economy, the government decides what goods and services will be produced and what prices will be charged for them. The government decides what methods of production will be used and how much workers will be paid. Many necessities like healthcare and education are provided for free. Currently, Cuba and North Korea have command economies.

Figure 4: A typical market. (Credit: Mabel, 2007)

Although command economies have a very centralized structure for economic decisions, market economies have a very decentralized structure. A **market** is an institution that brings together buyers and sellers of goods or services, who may be either individuals or businesses. The fresh vegetable market shown above is a good example of a market in which buyers and sellers are brought together. In a **market economy**, decision-making is decentralized. Market economies are based on **private enterprise**: the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals. Businesses supply goods and services based on demand. (In a command economy, by contrast, resources and businesses are owned by the government.) In a market economic system what goods and services are supplied depends on what is demanded. A person's income is based on his or her ability to convert resources (especially labor) into something that society values. The more society values the person's output, the higher the income (think South African celebrities such soccer player Itumeleng Khune and musician Mafikizolo). In this scenario, economic decisions are determined by market forces, not governments.

Most economies in the real world are mixed; they combine elements of command and market (and even traditional) systems. The South African

economy is positioned toward the market-oriented end of the spectrum. Many countries in Europe and Latin America (as in the case of South Africa), while primarily market-oriented, have a greater degree of government involvement in economic decisions than do countries such as the USA and Hong Kong, for example. China and Russia, while they are closer to having a market-oriented system now than several decades ago, remain closer to the command economy end of the spectrum. A rich resource of information about countries and their economies can be found on the Heritage Foundation's website (The Heritage Foundation: 2016), as the following Clear It Up feature discusses.

Note:

Which countries are considered economically free?

Who is in control of economic decisions? Are people free to do what they want and to work where they want? Are businesses free to produce when they want and what they choose, and to hire and fire as they wish? Are banks free to choose who will receive loans? Or does the government control these kinds of choices? Each year, researchers at the Heritage Foundation and the *Wall Street Journal* look at 50 different categories of economic freedom for countries around the world. They give each nation a score based on the extent of economic freedom in each category.

The Heritage Foundation's 2016 Index of Economic Freedom report ranked 178 countries around the world: some examples of the most free and the least free countries are listed in Table 1. Several countries were not ranked because of extreme instability that made judgments about economic freedom impossible. These countries include Afghanistan, Iraq, Syria, Somalia and Sudan.

The assigned rankings are based on estimates, yet even these rough measures can be useful for spotting trends. In 2015, 101 of the 178 included countries shifted toward greater economic freedom, although 77 of the countries shifted toward less economic freedom. South Africa is ranked as a "moderately" free market economy at number 80 out of 178 countries surveyed in 2016. In recent decades, the overall trend has been towards a *higher level of economic freedom around the world*.

Most Economic Freedom	Least Economic Freedom
1. Hong Kong	167. Timor-Leste
2. Singapore	168. Central African Republic
3. New Zealand	169. Argentina
4. Switzerland	170. Equatorial Guinea
5. Australia	171. Iran
6. Canada	172. Republic of the Congo
7. Chile	173. Eritrea
8. Ireland	174. Turkmenistan
9. Estonia	175. Zimbabwe
10. United Kingdom	176. Venezuela
11. United States of America	177. Cuba
12. Denmark	178. North Korea

Economic Freedoms, 2016(Source: The Heritage Foundation, 2016 Index of Economic Freedom, Country Rankings)

In real life there are no strictly traditional, command or absolutely free market economies. In the modern era, all economies are a mix of these different methods of economic organisation. So in most societies and countries today we find small pockets of traditional economic activity such as subsistence farming, hunting and small-scale home industries. However,

most economic activity is channelled through markets which, in some cases, might be regulated by government (command style). So what does a typical, modern mixed market economy look like? If we take the basic 2-sector circular flow model of economic activity and add a bit more detail to it, we should get a good idea.

The Full Circular Flow Model of Economic Activity

In the full circular flow model of economic activity, government is integrated into the real and monetary flows that accompany economic activity. We observe from the diagram below that both households (consumers) and government purchase the produce of firms. Government purchases in the product market would typically comprise items required to run offices and afford service delivery. These could include fuel, vehicles, uniforms, stationery, computers and so forth. We notice also that both firms and government purchase resources from households. Just like firms, government entities require staff (labor), accommodation (land) and equipment (capital) to operate effectively. These would be purchased in the resource market. The cost of these inputs to firms and government represent income to households who own these resources: wages for labor, rent for land, interest payments for capital and normal profit for entrepreneurship (in the case of firm owners).

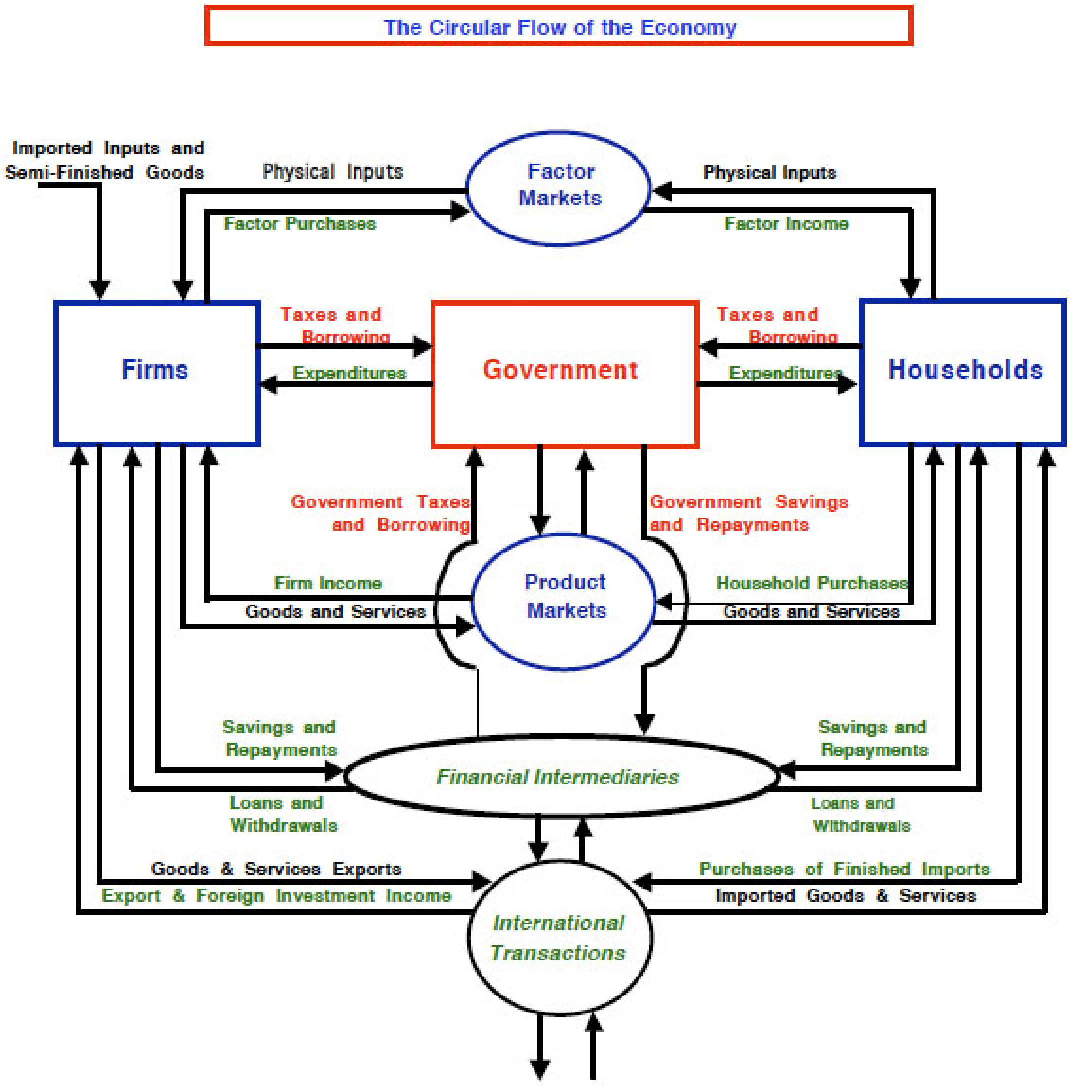


Figure 5. The full circular flow model of economic activity. (Credit: Circular Flow Diagram:2012)

The expenditures by government on both firms and households include the provision of public goods and services which, by their nature, cannot easily be traded in markets. These include infrastructure such as water provision, sewage reticulation, roads, traffic lights, street lights, parks, schools, clinics/hospitals, police services, prisons, public libraries and so forth. Government expenditures on households would also include transfer

payments such as pensions, subsidies and various types of grants to alleviate poverty. Government expenditure is financed primarily by corporate taxes on firms (including financial sector firms) and on households (income tax and VAT).

Economic activity (production, earning income from being engaged in production, and spending this income) in the modern era is no longer confined within the borders of countries. The world is increasingly interconnected and countries today have strong links with each other in the areas of trade and other spheres of cooperation. So we note from Figure 5 that not only do countries engage in importing and exporting goods and services but also that investment funds flow into, as well as out of, countries.

Of course modern market economies, based as they are on trade, could not exist without the amazing invention of money and the financial sector firms that manage it (banks and other financial intermediaries). Virtually all the income in an economy is banked at some point. Most expenditure is funded from bank accounts. Most of the credit that drives spending is advanced by banks and other similar financial intermediaries. So we notice from Figure 4 that households, firms and government all have bank accounts (and save) and also borrow money from the banks.

The basic picture of the different elements of a modern economy provided by the circular flow model is very useful for tracking the sources of problems societies often face such as unemployment and inflation. Let's take unemployment which is a very serious problem for South Africa. In the second quarter (April-June) of 2015, the official measure of unemployment was 25% which means that one in four people who were willing and able to work could not (Statistics South Africa: 2015). Unemployment can often be traced to a lack of economic activity i.e. production, spending and earning income. It follows that, if production declines, then unemployment will increase. But production will generally only decline if sales decline. Sales decline mostly because consumers spend less and spending is likely to decline if income falls. No wonder it is called the circular flow of economic activity!

So it seems that the level of economic activity will fluctuate with the amount of spending that takes place in an economy because spending determines sales and production. However, we need to notice that someone's spending is always someone else's income: If I spend R100 at Spar, the owner of the shop earns R100 worth of income. **Injections** of income/spending into an economy boost production while **leakages** of income/spending lead to declining production. The following injections of spending/income into the domestic economy can be identified in the circular flow model: Consumption spending by domestic households (C), Investment spending by domestic firms and government (I), Government consumption spending (G) and Export spending by foreigners on domestic produce (X). Leakages of spending/income from the domestic economy include domestic Savings (S), Taxes (T) and Import spending (Z).

If the circular flow model is good for identifying the possible sources of economic problems such as declining levels of production (low or declining economic growth), then it is equally good for suggesting ways in which spending in the economy may be managed or manipulated to achieve higher or lower levels of economic activity. Notice government's central position in the circular flow of economic activity. It can easily increase total spending in the economy by reducing taxes on households and firms. It can just as easily depress total spending by increasing taxes. Government can also change its own level of spending in line with economic policy. When government adjusts taxes and/or its own spending in line with bigger economic goals this is called "fiscal policy".

The level of economic activity is directly correlated with the amount of money in circulation in the circular flow of economic activity. More spending, production and income requires more money to be created to match the required level of economic activity. This is where the financial sector in a modern economy becomes very important since it is this group of organisations/firms that creates an economy's money. We will see how this happens in later chapters. Increasing or decreasing the amount of money in the economy to achieve certain economic goals is termed "monetary policy".

Regulations: The Rules of the Game

Markets are often regulated by government. As we have noted, there is no such thing as an absolutely free market. Regulations always define the "rules of the game" in the economy. Economies that are primarily market-oriented have fewer regulations—ideally just enough to maintain an even playing field for participants. At a minimum, these laws govern matters like safeguarding private property against theft, protecting people from violence, enforcing legal contracts, preventing fraud, and collecting taxes. Conversely, even the most command-oriented economies operate using markets. How else would buying and selling occur? But the decisions of what will be produced and what prices will be charged are heavily regulated. Heavily regulated economies often have **underground economies**, which are markets where the buyers and sellers make transactions without the government's approval.

The question of how to organize economic institutions is typically not a clear-cut choice between all market or all government, but instead involves a balancing act over the appropriate combination of market freedom and government rules. Hence most economies are mixed market economies (i.e. a mix of traditional, command and mainly market).

Figure 6: Cargo ships are one mode of transportation for shipping goods in the global economy. (Credit: Raul Valdez/Flickr Creative Commons)

The Rise of Globalization

Recent decades have seen a trend toward **globalization**, which is the expanding cultural, political, and economic connections between people around the world. One measure of this is the increased buying and selling of goods, services, and assets across national borders—in other words, international trade and financial capital flows.

Globalization has occurred for a number of reasons. Improvements in shipping, as illustrated by the container ship shown in Figure 6, and air cargo have driven down transportation costs. Innovations in computing and telecommunications have made it easier and cheaper to manage long-distance economic connections of production and sales. Many valuable products and services in the modern economy can take the form of information—for example: computer software and applications; financial advice; travel planning; music, books and movies; and blueprints for designing a building. These products and many others can be transported over telephones and computer networks at ever-lower costs. Finally, international agreements and treaties between countries have encouraged greater trade.

Table 2 presents one measure of globalization. It shows the percentage of domestic economic production that was exported for a selection of countries from 2010 to 2013, according to The World Bank. **Exports** are the goods and services that are produced domestically and sold abroad. **Imports** are the goods and services that are produced abroad and then sold domestically. The size of total production in an economy is measured by the **gross domestic product (GDP)**. Thus, the ratio of exports divided by GDP measures what share of a country's total economic production is sold in other countries.

Country	2010	2011	2012	2013
Higher Income Countries				
United States	12.4	13.6	13.6	13.5
Belgium	76.2	81.4	82.2	82.8
Canada	29.1	30.7	30.0	30.1
France	26.0	27.8	28.1	28.3
Middle Income Countries				
Brazil	10.9	11.9	12.6	12.6
South Africa	28.6	30.4	29.7	31.0
South Korea	49.4	55.7	56.3	53.9
Lower Income Countries				
Chad	36.8	38.9	36.9	32.2
China	29.4	28.5	27.3	26.4
India	22.0	23.9	24.0	24.8
Nigeria	25.3	31.3	31.4	18.0

The Extent of GlobalizationExports of goods and services (% of GDP) (Source: World Bank: 2016)

In recent decades, the export/GDP ratio has generally risen, both worldwide and for the South African economy. South Africa's share of exports as a percent of GDP is similar to some other high and low income countries. In general smaller economies like those of South Africa, Belgium, Korea, and

Canada need to trade across their borders with other countries to take full advantage of division of labor, specialization, and economies of scale.

Table 2 also shows that many medium and low income countries around the world, like South Africa and China, have also experienced a surge of globalization in recent decades. If an astronaut in orbit could put on special glasses that make all economic transactions visible as brightly colored lines and look down at Earth, the astronaut would see the planet covered with connections.

Note:

Decisions ... Decisions in the Social Media Age

The world we live in today provides nearly instant access to a wealth of information via the internet. Search engines perform trillions of searches each year. Some relatively new information forums, such as Facebook,Twitter and WhatsApp are rapidly changing how information is distributed; hence, influencing decision making. In 2014, the Pew Research Center reported that 71% of online adults use Facebook. Facebook post topics range from politics to sport to celebrity singers and performers, to farming.

Information helps us make decisions. Decisions as simple as what to wear today, or, how many reporters should be sent to cover a crash. Each of these decisions is an economic decision. After all, resources are scarce. If ten reporters are sent to cover an accident, they are not available to cover other stories or complete other tasks. Information provides the knowledge needed to make the best possible decisions on how to utilize scarce resources. Welcome to the world of economics!

Key Concepts and Summary

Societies can be organized as traditional, command, or market-oriented economies. Most societies are a mix. The last few decades have seen globalization evolve as a result of growth in commercial and financial

networks that cross national borders, making businesses and workers from different economies increasingly interdependent.

Self-Check Questions

Exercise:

Problem:

The chapter defines *private enterprise* as a characteristic of market-oriented economies. What would *public enterprise* be? *Hint*: It is a characteristic of command economies.

Solution:

Public enterprise means the factors of production (resources and businesses) are owned and operated by the government.

Review Questions

Exercise:

Problem:

What are the three ways that societies can organize themselves economically?

Solution:

According to tradition, command or a market system

Exercise:

Problem:

What is globalization? How do you think it might have affected the economy over the past decade?

Solution:

Offer your solution here...

Critical Thinking Questions

Exercise:

Problem:

Why do you think that most modern countries’ economies are a mix of command and market types?

Exercise:

Problem:

Can you think of ways that globalization has helped you economically? Can you think of ways that it has not?

References

Circular Flow Diagram. [Figure]. Wikispaces, 15 April 2012. Available: https://econ101-powers-sectione.wikispaces.com/Circular+Flow+Diagram. (Accessed: 17 Oct 2016)

Mabel, J. Ballard Farmers’ Market - Vegetables. Photograph, July 29, 2007. Available: https://commons.wikimedia.org/wiki/File:Ballard_Farmers’_Market_-_vegetables.jpg.(Accessed 21 Oct 2016)

Palmer, N. A Kenyan Farmer at Work in the Mount Kenya Region. Photograph, October 18, 2010. Available: https://en.wikipedia.org/wiki/File:2DU_Kenya_86_%285367322642%29.jpg.(Accessed 21 October 2016)

Pew Research Center. 2015. Social Networking Fact Sheet. Available: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ (Accessed: 17 Oct 2016)

Statistics South Africa. 2015. Quarterly Labour Force Survey, Quarter 2: 2015. Statistical release, P0211. Pretoria: Statistics South Africa. Available: http://www.statssa.gov.za/publications/P0211/P02112ndQuarter2015.pdf. (Accessed: 17 Oct 2016)

The Heritage Foundation. 2016. 2016 Index of Economic Freedom. Available: http://www.heritage.org/index/ranking (Accessed: 17 Oct 2016)

World Bank. 2016. World Data Bank. World Development Indicators. Exports of goods and services (% of GDP). Available: http://databank.worldbank.org/data/ (Accessed: 17 Oct 2016)

Glossary

command economy
: an economy where economic decisions are passed down from government authority and where resources are owned by the government

exports
: products (goods and services) made domestically and sold abroad

globalization
: the trend in which buying and selling in markets have increasingly crossed national borders

gross domestic product (GDP)
: measure of the size of total production in an economy

imports
: products (goods and services) made abroad and then sold domestically

market
: interaction between potential buyers and sellers; a combination of demand and supply

market economy

an economy where economic decisions are decentralized, resources are owned by private individuals, and businesses supply goods and services based on demand

private enterprise
: system where the means of production (resources and businesses) are owned and operated by private individuals or groups of private individuals

traditional economy
: typically an agricultural economy where things are done the same as they have always been done

underground economy
: a market where the buyers and sellers make transactions in violation of one or more government regulations

Introduction to Choice in a World of Scarcity
class="introduction"

Figure 1: In general, the higher the degree, the higher the salary. So why aren't more people pursuing higher degrees? The short answer: choices and tradeoffs. (Credit: Creative Commons)

Note:
Choices ... To What Degree?
So economics is really just about how and why people make the choices they do? Well let's see if this applies to your choice and decision to enroll for higher education studies. Three or four more years of study beyond Grade 12 to earn a diploma or degree seems like a quite a lot of time to be out of the labor market when you could be working and earning an income. In making the choice to study further, you must have wondered whether it would be worthwhile giving up the alternative possibility of a job. Well let's check out some starting salaries for new graduates in South Africa. According to the web article *The degrees which will get you the highest starting salary in South Africa* (2015), a BSc Engineering will start at R19 069, Bachelor of Education (teacher)- R15 825, BTech - R13 008, BCom

(Accounting) - R11 070, Matric - R5 830. From a financial perspective, it therefore seems worthwhile studying beyond school.
Given these statistics, we might expect a lot of people to choose to go to college or university and at least earn a diploma or a degree. In fact the number of graduates in the South African labor market grew from 463 000 in 1995 to 1,1 million in 2011 (Altbeker and Storme: 2013). Altbeker and Storme (2013) confirm that the rate of graduate unemployment in South Africa is less than 5%, for matrics unemployment is about 29% and for those without a matric qualification unemployment is about 42%. Why then do only 18% of South African school leavers enroll at universities? (Nkosi: 2015) Why do about half of these new enrolling students drop out of university study in their first year? Why do only about 15% of South African students who enroll at university eventually graduate? (Mtshali: 2013) How many of these facts are due to choices?
This brings us to the subject of this chapter: why people make the choices they make and how economists go about explaining those choices.

Note:
Introduction to Choice in a World of Scarcity
In this chapter, you will learn about:

- How Individuals Make Choices Based on Their Budget Constraint
- The Production Possibilities Frontier and Social Choices

You will learn quickly when you examine the relationship between economics and scarcity that choices involve tradeoffs. Every choice has a cost.

In 1968, the Rolling Stones recorded the song “You Can’t Always Get What You Want.” Economists chuckled, because they had been singing a similar tune for decades. English economist Lionel Robbins (1898–1984), in his *Essay on the Nature and Significance of Economic Science* in 1932, described not always getting what you want in this way:

"The time at our disposal is limited. There are only twenty-four hours in the day. We have to choose between the different uses to which they may be put. ... Everywhere we turn, if we choose one thing we must relinquish (give up) others which, in different circumstances, we would wish not to have relinquished. Scarcity of means to satisfy given ends is an almost ubiquitous (common) condition of human nature."

Because people live in a world of scarcity, they cannot have all the time, money, possessions, and experiences they wish. Neither can society.

This chapter will continue our discussion of the economic way of thinking by introducing some critical economic concepts: scarcity of resources (or factors of production), choice, opportunity cost, marginal decision making and diminishing returns, the law of increasing opportunity cost and the concept of efficiency. Later, it will consider whether the economic way of thinking accurately describes either how choices *are* made or how they *should* be made.

REFERENCES

Anon. Spelman Graduates. Photograph, Undated. http://fierceforblackwomen.com/wp-content/uploads/2015/10/Spelman-graduates-Creative-Commons.jpg.(Accessed: 22 Nov 2016)

Altbeker, A. and Storme, E. 2013. Graduate unemployment in South Africa: a much exaggerated problem. Johannesburg: Centre for Development and Enterprise. CDE Insight. Available: http://sun025.sun.ac.za/portal/page/portal/Arts/Departments/political_science/docs/Tab/GRADUATE_UNEMPLOYMENT_IN_SOUTH_AFRICA__A_much_exaggerate.pdf (Accessed: 22 Apr 2016)

The degrees which will get you the highest starting salary in South Africa. 2015. Mybroadband, 6 Sep. Available: http://mybroadband.co.za/news/business/137306-the-degrees-which-will-get-you-the-highest-starting-salary-in-south-africa.html (Accessed: 22 Apr 2016)

Mtshali, N. 2013. Only 15% of SA university students graduate. IOL News, 13 June. Available: http://www.iol.co.za/lifestyle/family/kids/only-15-of-sa-university-students-graduate-1531809 (Accessed: 22 Apr 2016)

Nkosi, B. 2015. Of the 18% of matrics registered at university, half drop out. Mail and Guardian, 21 May. Available: http://mg.co.za/article/2015-05-21-18-of-matrics-register-at-universities-half-drop-out (Accessed: 22 Apr 2016)

Robbins, L.C. 1932. Essay on the nature and significance of economic science. London: Macmillan. Available: https://mises.org/library/essay-nature-and-significance-economic-science (Accessed: 22 Apr 2016)

How Individuals Make Choices Based on Their Budget Constraint

By the end of this section, you will be able to:

- Calculate and graph budgets constraints
- Explain opportunity sets and opportunity costs
- Evaluate the law of diminishing marginal utility
- Explain how marginal analysis and utility influence choices

Figure 1. Burgers or bus tickets?

Consumers have a limited amount of income to spend on the things they need and want. People have to budget. Suppose Sipho has R100 in spending money each week that he can allocate between bus tickets for getting to university and the burgers that he eats for lunch. Burgers cost R20 each, and bus tickets are R5 each from the city center . Figure 2 shows Sipho's **budget constraint**, that is, the outer boundary of his **opportunity set**. The opportunity set identifies all the opportunities for spending within his budget. The budget constraint indicates all the combinations of burgers and bus tickets Sipho can afford when he spends his entire budget, given the prices of the two goods.

The Budget Constraint: Sipho's Consumption Choice Opportunity Frontier

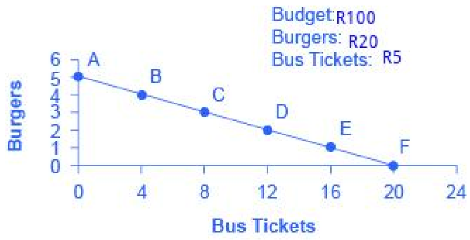


Figure 2: Each point on the budget constraint represents a combination of burgers and bus tickets whose total cost adds up to Sipho's budget of R100. The slope of the budget constraint is determined by the relative price of burgers and bus tickets. All along the budget set, giving up one burger means gaining four bus tickets.

The vertical axis in the figure shows burger purchases and the horizontal axis shows bus ticket purchases. If Sipho spends all his money on burgers, he can afford five per week. (R100 per week/R20 per burger = 5 burgers per week.) But if he does this, he will not be able to afford any bus tickets. This choice (zero bus tickets and five burgers) is shown by point A in the figure. Alternatively, if Sipho spends all his money on bus tickets, he can afford 20 per week. (R100 per week/R5 per bus ticket = 20 bus tickets per week.) Then, however, he will not be able to afford any burgers. This alternative choice (20 bus tickets and zero burgers) is shown by point F.

If Sipho is like most people, he will choose some combination that includes both bus tickets and burgers. That is, he will choose some combination on the budget constraint that connects points A and F. Every point on (or inside) the constraint shows a combination of burgers and bus tickets that Sipho can afford. Any point outside the constraint is not affordable, because it would cost more money than Sipho has in his budget.

The budget constraint clearly shows the tradeoff Sipho faces in choosing between burgers and bus tickets. Suppose he is currently at point D, where he can afford 12 bus tickets and two burgers. What would it cost Sipho for one more burger? It would be natural to answer R20, but that's not the way economists think. Instead they ask, how many bus tickets would Sipho have to give up to get one more burger, while staying within his budget? The answer is four bus tickets. That is the true cost to Sipho of one more burger.

The Concept of Opportunity Cost

Economists use the term **opportunity cost** to indicate what must be given up to obtain something that is desired. The idea behind opportunity cost is that the cost of one item is the lost opportunity to do or consume something else; in short, opportunity cost is the value of the next best alternative foregone (not chosen). For Sipho, the opportunity cost of another burger is the four bus tickets he would have to give up. He would decide whether or not to choose the burger depending on whether the value of the burger exceeds the value of the forgone alternative—in this case, bus tickets. Since people must choose, they inevitably face tradeoffs in which they have to give up things they desire to get other things they desire more.

Note:
View this website for an example of opportunity cost—paying someone else to wait in line for you. This business figured out that many people cannot afford to waste time standing in queues (that the opportunity cost for them of doing this was just too great). They would rather incur the smaller opportunity cost of paying someone else stand in the queue to get what they need (ID, passport etc.)

A fundamental principle of economics is that every choice has an opportunity cost. If you sleep through your economics class (not recommended, by the way), the opportunity cost is the learning you miss from not attending class. If you spend your income on video games, you cannot spend it on movies. If you choose to marry one person, you give up the opportunity to marry anyone else. In short, opportunity cost is all around us and part of human existence.

Note:
Understanding Budget Constraints
Budget constraints are easy to understand.
Step 1: The equation for any budget constraint is:
Equation:

$$\text{Budget} = \text{P}_1 \times \text{Q}_1 + \text{P}_2 \times \text{Q}_2$$

where P and Q are the price and quantity of items purchased and Budget is the amount of income one has to spend.
Step 2. Apply the budget constraint equation to the scenario. In Sipho’s case, this works out to be:
Equation:

$$\begin{aligned} \text{Budget} &= \text{P}_1 \times \text{Q}_1 + \text{P}_2 \times \text{Q}_2 \\ \text{R100 budget} &= \text{R20 per burger} \times \text{quantity of burgers} + \text{R5 per bus ticket} \times \text{quantity of bus tickets} \\ \text{R100} &= \text{R20} \times \text{Q}_{\text{burgers}} + \text{R5} \times \text{Q}_{\text{bus tickets}} \end{aligned}$$

Step 3. Using a little algebra, we can turn this into the familiar equation of a straight line:
Equation:

$$\text{y} = \text{b} + \text{mx}$$

Where: y = the y-axis variable, b = the value of the y-axis intercept, m = gradient of the budget line/constraint, x = the x-axis variable.
For Sipho, this is:
R100 = R20 x Qburgers + R5 x Qbus tickets
Step 4. Next we get the budget constraint equation into straight line format by solving for the y-variable (burgers). We do this by dividing both sides by R5
R100/R5 = (R20/4) x burgers + (R5/R5) x bus tickets
20 = 4 x burgers + 1 x bus tickets. Now rearranging and solving for burgers (y-variable)
4 x burgers = 20 - 1 x bus ticket. Now dividing both sides by 4
burgers = 5 - (1/4) x bus tickets
Step 5. Notice that this equation fits the budget constraint in Figure 1. The vertical intercept is 5 and the slope is –0.25 or -1/4, just as the equation says. If you plug 20 bus tickets into the equation, you get 0 burgers. If you plug other numbers of bus tickets into the equation, you get the results shown in Table 1, which are the points on Sipho’s budget constraint.

Point	Quantity of Burgers (at R20)	Quantity of Bus Tickets (at R5)
A	5	0
B	4	4
C	3	8
D	2	12
E	1	16
F	0	20

Step 6. Notice that the slope of a budget constraint always shows the opportunity cost of the good which is on the horizontal axis. For Sipho, the slope is −0.25, indicating that for every four bus tickets he buys, Sipho must give up 1 burger.
There are two important observations here. First, the algebraic sign of the slope is negative, which means that the only way to get more of one good is to give up some of the other. Second, the slope is defined as the price of bus tickets (whatever is on the horizontal axis in the graph) divided by the price of burgers (whatever is on the vertical axis), in this case R5/R20 = 0.25. So if you want to determine the opportunity cost quickly, just divide the two prices.

Identifying Opportunity Cost

In many cases, it makes sense to refer to the opportunity cost as the price. If your cousin buys a new bicycle for R3 500, then R3 500 measures the amount of "other consumption" that he has given up. For practical purposes, there may be no special need to identify the specific alternative product or products that could have been bought with that R3 500, but sometimes the price as measured in Rands (or any other currency) may not accurately capture the true opportunity cost. This is especially the case when costs of time are involved.

For example, consider a boss who decides that all employees will attend a two-day workshop to "build team spirit." The monetary cost of the event may involve hiring an outside consulting firm to run the workshop, as well as accommodation for all participants. But an opportunity cost for the employer exists as well: during the two days of the workshop, none of the employees are doing any other work.

Attending college or university is another case where the opportunity cost exceeds the monetary cost. The monetary costs of attending college include tuition, books, room and board, and other expenses. But in addition, during the hours that you are attending class and studying, it is impossible to work at a paying job. Thus, college imposes both an out-of-pocket cost and an opportunity cost of lost earnings.

In some cases, realizing the opportunity cost can alter behavior. Imagine, for example, that you spend R30 on lunch every day at university. You may know perfectly well that bringing a lunch from home would cost only R15 a day, so the opportunity cost of buying lunch at the university restaurant is R15 each day (that is, the R30 buying lunch costs minus the R15 your lunch from home would cost). R15 opportunity cost each day does not seem to be that much. However, if you project what that adds up to in a year (250 days a year × R15 per day = R3 750) it could be the cost, perhaps, of a good holiday or a flatscreen TV. If the opportunity cost of buying lunch at the university is described as "a nice vacation or a flatscreen TV" instead of "R15 a day," you might make different choices.

Marginal Decision-Making and Diminishing Marginal Utility

The budget constraint idea helps to emphasize that most choices in the real world are not about getting all of one thing or all of another; that is, they are not about choosing either the point at one end of the budget constraint or else the point all the way at the other end. Instead, most choices involve **marginal analysis**, which means comparing the benefits and costs of choosing a little more or a little less of a good. Marginal analysis is very important in the economic study of how people make choices.

People desire goods and services for the satisfaction or **utility** those goods and services provide. Utility (satisfaction) is subjective but that does not make it less real. Economists usually assume that the more of some good one consumes (for example, wors rolls), the more utility one obtains. At the same time, the utility a person receives from consuming the first unit of a good is typically more than the utility received from consuming the fifth or the tenth unit of that same good (after all, you can get sick of it!). When Sipho chooses between burgers and bus tickets, for example, the first few bus rides that he chooses might provide him with a great deal of utility—perhaps they help him get to a job interview or a doctor's appointment. But later bus rides might provide much less utility—they may only serve to kill time on a rainy day. Similarly, the first burger that Sipho chooses to buy may be on a day when he missed breakfast and is very hungry. However, if Sipho has a burger every single day, the last few burgers in the week may taste pretty boring. The general pattern that consumption of the first few units of any good tends to bring a higher level of utility (satisfaction) to a person than consumption of later units is a common pattern. Economists refer to this pattern as the **law of diminishing marginal utility**, which means that as a person receives more of a good, the additional (or marginal) utility or satisfaction from each additional unit of the good consumed declines. In other words, the first wors roll brings more satisfaction than the sixth (which might actually make you sick!).

The law of diminishing marginal utility explains why people and societies do not often make all-or-nothing choices. You would not say, "My favorite food is ice cream, so I will eat nothing but ice cream from now on." Instead, even if you get a very high level of utility from your favorite food, if you ate only that, the additional or marginal utility from those last few servings of ice cream would not be very high. Similarly, most workers do not say: "I enjoy leisure, so I'll never work." Instead, workers recognize that even though some leisure is very nice, a combination of all leisure and no income is not so attractive. The budget constraint model suggests that when people make choices in a world of scarcity, they will use marginal analysis and think about whether they would prefer a little more or a little less.

Sunk Costs

In the budget constraint model, all decisions involve what will happen next: that is, what quantities of goods will you consume, how many hours will you work, or how much will you save. These decisions do not look back to past choices. Thus, the budget constraint idea assumes that **sunk costs**, which are costs that were incurred in the past and cannot be recovered, should not affect the current decision or choice. This is a very important point if we are to identify and understand the patterns of choices that people make. A knowledge of sunk costs is very important for accurate economic analysis.

Let's take the case of Selena, who pays R60 to see a movie, but after watching the film for 30 minutes, she knows that it is truly terrible. Should she stay and watch the rest of the movie because she paid for the ticket, or should she leave? The money she spent buying the ticket is a sunk cost, and unless the theater manager is feeling kindly, Selena will not get a refund. But staying in the movie still means paying an opportunity cost in time. Her choice is whether to spend the next 90 minutes watching the movie and suffering from boredom or to walk out and do something else. The lesson of sunk costs is to forget about the money and time that has been lost and instead to focus on the marginal (additional or extra) costs and benefits of current and future options.

For consumers and firms alike, dealing with sunk costs can be frustrating. It often means admitting an earlier error in judgment and then moving on. Many firms, for example, find it hard to give up on a new product that is doing poorly because they spent so much money in creating and launching the product. But the lesson of sunk costs is to ignore them and make decisions based on what will happen in the future. In fact this is a life lesson: do not live in the past. We should not let the past dictate our behavior, or the choices and decisions we make. It is the future that counts. We need to weigh up the benefits and costs of future actions.

Take a smoker, for example. She started by experimenting with just one cigarette. She liked it and so tried a few more which she bought at the spaza shop. The extra/additional/marginal cigarettes resulted in additional/extra/marginal monetary cost for her. However the additional cigarettes also gave her additional/extra/marginal satisfaction (benefit). Which, unfortunately for her long term health, is why our young friend is today a smoker. Now let's say she decides to stop smoking for health reasons. The reason for this decision would surely be that she has calculated that the future benefits (better health, less medical bills) of quitting smoking (from now on) outweigh the future costs (of future cigarettes consumed plus medical bills). Her decision to quit smoking now would not have been influenced by how much she had already spent on smoking. That is in the past.

From a Model with Two Goods to One of Many Goods

The budget constraint diagram containing just two goods (burgers and bus tickets), like most models used in this book, this is not realistic (it is a simplification of real life). After all, in a real modern economy people choose from thousands of goods and services. However, thinking about a model with many goods is a straightforward extension of what we discussed here. The graph with two goods that was presented here clearly illustrates that every choice has an opportunity cost, which is the point that does carry over to the real world.

Key Concepts and Summary

Economists see the real world as one of scarcity: that is, a world in which people’s desires exceed what is possible. As a result, economic behavior involves tradeoffs in which individuals, firms, and society must give up something that they desire to obtain things that they desire more. Individuals face the tradeoff of what quantities of goods and services to consume. The budget constraint, which is the frontier of the opportunity set, illustrates the range of choices available. The slope of the budget constraint is determined by the relative price of the choices. Choices beyond the budget constraint are not affordable.

Opportunity cost measures cost by what is given up in exchange. Sometimes opportunity cost can be measured in money, but it is often useful to consider time as well, or to measure it in terms of the actual resources that must be given up.

Most economic decisions and tradeoffs are not all-or-nothing. Instead, they involve marginal analysis, which means they are about decisions on the margin, involving a little more or a little less. The law of diminishing marginal utility points out that as a person receives more of something—whether it is a specific good or another resource—the additional marginal gains tend to become smaller. Because sunk costs occurred in the past and cannot be recovered, they should be disregarded in making current decisions.

Self-Check Questions

Exercise:

Problem:

Suppose Sipho's town raised the price of bus tickets to R10 per trip (while the price of burgers stayed at R20 and his budget remained R100 per week.) Draw Sipho's new budget constraint. What happens to the opportunity cost of bus tickets?

Solution:

The opportunity cost of bus tickets is the number of burgers that must be given up to obtain one more bus ticket. Originally, when the price of bus tickets was R5 per trip, this opportunity cost was 5/20 = .25 burgers. The reason for this is that at the original prices, one burger (R20) costs the same as four bus tickets (R5), so the opportunity cost of a burger is four bus tickets, and the opportunity cost of a bus ticket is .25 (one quarter) of a burger. With the new, higher price of bus tickets, the opportunity cost of bus tickets increases to R10/R20 or 0.50. You can see this graphically in Figure 3 below since the slope of the new budget constraint is steeper than the original one. If Sipho spends all of his budget on burgers, the higher price of bus tickets has no impact so the vertical intercept of the budget constraint (5 burgers) is the same. If he spends all of his budget on bus tickets, he can now afford only half as many, so the new horizontal intercept (10 bus tickets) is half as much.

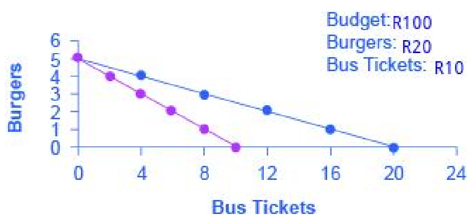


Figure 3.

Review Questions

Exercise:

Problem:

Explain why scarcity leads to tradeoffs (forced to make choices which automatically results in opportunity cost).

Exercise:

Problem:

Explain why individuals make choices that are directly on the budget constraint, rather than inside the budget constraint or outside it.

Critical Thinking Question

Exercise:

Problem:

Suppose Sipho’s town raises the price of bus tickets from R5 to R10 and the price of burgers rises from R20 to R40. Why is the opportunity cost of bus tickets unchanged? Suppose Sipho’s weekly spending money increases from R100 to R200. How is his budget constraint affected from all three changes? Explain.

Problems

Smangele has a weekly budget of R240, which she likes to spend on magazines and eating out at her favorite restaurant. Use this information to answer the following 4 questions:

Exercise:

Problem:

If the price of a magazine is R40 each, what is the maximum number of magazines she could buy in a week?

Exercise:

Problem:

If the price of her favourite restaurant meal is R120, what is the maximum number of times she can eat out in a week?

Exercise:

Problem:

Draw Smangele's budget constraint with restaurant meals on the horizontal axis and magazines on the vertical axis. What is the slope of the budget constraint?

Exercise:

Problem: What is Smangele's opportunity cost of purchasing her favourite restaurant meal?

Glossary

budget constraint
: all possible consumption combinations of goods that someone can afford, given the prices of goods, when all income is spent; the boundary of the opportunity set

law of diminishing marginal utility
: as we consume more of a good or service, the utility we get from additional units of the good or service tend to become smaller than what we received from earlier units

marginal analysis
: examination of decisions on the margin, meaning a little more or a little less from the status quo

opportunity cost
: measures cost by what is given up in exchange; opportunity cost measures the value of the forgone alternative

opportunity set

all possible combinations of consumption that someone can afford given the prices of goods and the individual’s income

sunk costs
: costs that are made in the past and cannot be recovered

utility
: satisfaction, usefulness, or value one obtains from consuming goods and services

The Production Possibilities Frontier and Social Choices

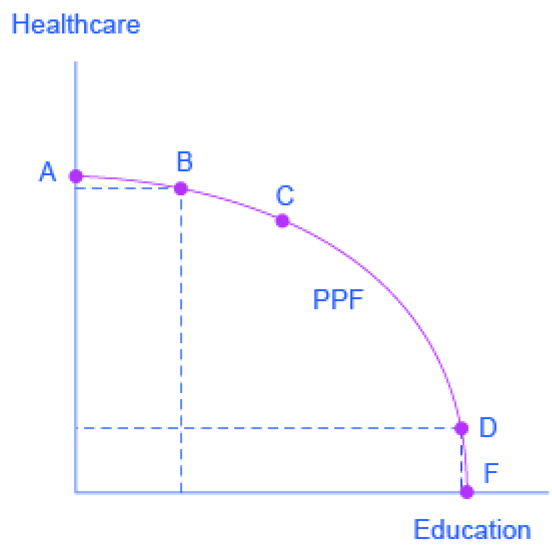
By the end of this section, you will be able to:

- Interpret production possibilities frontier graphs
- Contrast a budget constraint and a production possibilities frontier
- Explain the relationship between a production possibilities frontier and the law of diminishing returns
- Contrast productive efficiency and allocative efficiency
- Define comparative advantage

Just as individuals cannot have everything they want and must instead make choices, society as a whole cannot have everything it might want, either. This section of the chapter will explain the constraints faced by society, using a model called the **production possibilities curve or frontier (PPC/PPF)**. The PPC illustrates all possible combinations of maximum and efficient production of two goods/services given available resources and technology at a given time. While the budget constraint model may be used to explain choices made by an individual, the PPC can be used to explain the choices made by groups of individuals (communities or societies). There are more similarities than differences between individual choice and social (communal) choice. As you read this section, focus on the similarities.

Because society has limited resources (e.g., labor, land, capital and entrepreneurial talent/skill) at any point in time, there is a limit to the quantities of goods and services it can produce. Suppose a society desires two services, healthcare and education. This situation is illustrated by the production possibilities frontier in the figure below.

A Healthcare vs. Education Production Possibilities Curve/Frontier



This production possibilities frontier shows a tradeoff between devoting society's resources to healthcare and devoting them to education. At A all resources go to healthcare and at B, most go to healthcare. At D most resources go to education, and at F, all go to education.

In Figure 1 healthcare is shown on the vertical axis and education is shown on the horizontal axis. If the society were to allocate all of its resources to healthcare, it could produce at point A. But it would not have any resources to produce education. If it were to allocate all of its resources to education, it could produce at point F. Alternatively, the society could choose to produce any combination of healthcare and education shown on the production possibilities curve or frontier. In effect, the production possibilities curve/frontier plays the same role for society as the budget constraint plays for Sipho. Society can choose any combination of the two goods on or inside the PPF. But it does not have enough resources to produce outside the PPF.

Most important, the production possibilities frontier clearly shows the tradeoff between healthcare and education. Suppose society has chosen to

operate at point B, and it is considering producing more education. Because the PPF is downward sloping from left to right, the only way society can obtain more education is by giving up some healthcare. That is the tradeoff society faces. Suppose it considers moving from point B to point C. What would the opportunity cost be for the additional education? The opportunity cost would be the healthcare society has to give up. Just as with Sipho's budget constraint, the opportunity cost is shown by the slope of the production possibilities frontier.

Note:

What's the difference between a budget constraint and a PPF?

There is one major difference between a budget constraint and a production possibilities curve/frontier: A budget constraint is a straight line whereas the PPC/PPF has a curved shape because of the law of diminishing returns (also known as the law of increasing opportunity cost). With the PPC/PPF, as more resources are devoted to the production of a good or service, so the returns from these resources will diminish and the opportunity cost of additional production will increase.

Whether or not we have specific numbers, conceptually we can measure the opportunity cost of additional education as society moves from point B to point C on the PPF. The additional education is measured by the horizontal distance between B and C. The foregone healthcare is given by the vertical distance between B and C. The slope of the PPF between B and C is (approximately) the vertical distance (the "rise") over the horizontal distance (the "run"). This is the opportunity cost of the additional education.

The Shape of the PPF and the Law of Diminishing Returns

The budget constraints presented earlier in this chapter, showing individual choices about what quantities of goods to consume, were all straight lines. The reason for these straight lines was that the slope of the budget constraint was determined by relative prices of the two goods in the consumption budget constraint. However, the production possibilities

frontier for healthcare and education was drawn as a curved line. Why does the PPC/PPF have a different shape?

To understand why the PPC/PPF is curved, start by considering point A at the top left-hand side of the PPF. At point A, all available resources are devoted to healthcare and none are left for education. This situation would be extreme and even ridiculous. For example, children are seeing a doctor every day, whether they are sick or not, but not attending school. Now imagine that some of these resources are diverted from healthcare to education, so that the economy is at point B instead of point A. Diverting some resources away from A to B causes relatively little reduction in health because the last few marginal (extra or additional) Rands going into healthcare services are not producing much additional gain in health. However, putting those marginal Rands into education, which is completely without resources at point A, can produce relatively large gains. For this reason, the shape of the PPF from A to B is relatively flat, representing a relatively small drop-off in health and a relatively large gain in education.

Now consider the other end, at the lower right, of the production possibilities frontier. Imagine that society starts at choice D, which is devoting nearly all resources to education and very few to healthcare, and moves to point F, which is devoting *all* spending to education and none to healthcare. You can imagine that in the movement from D to F, the last few doctors must become high school science teachers, the last few nurses must become school librarians rather than dispensers of vaccinations, and the last few emergency rooms are turned into kindergartens. The gains to education from adding these last few resources to education are very small. However, the opportunity cost lost to health will be fairly large, and thus the slope of the PPF between D and F is steep, showing a large drop in health for only a small gain in education.

The lesson is not that society is likely to make an extreme choice like devoting no resources to education at point A or no resources to health at point F. Instead, the lesson is that the gains from committing additional marginal resources to education depend on how much is already being spent. If on the one hand, very few resources are currently committed to education, then an increase in resources used can bring relatively large

gains. On the other hand, if a large number of resources are already committed to education, then committing additional resources will bring relatively smaller gains.

This pattern is common enough that it has been given a name: the **law of diminishing returns**, (also known as the law of increasing opportunity cost) which holds that as additional amounts of resources are added to a certain purpose, the marginal (extra or additional) benefit from those additional increments will decline. When government spends a certain amount more on reducing crime, for example, the original gains in reducing crime could be relatively large. But additional increases typically cause relatively smaller reductions in crime, and paying for enough police and security to reduce crime to nothing at all would be tremendously expensive.

The outward curvature of the production possibilities curve/frontier shows that as additional resources are added to education, moving from left to right along the horizontal axis, the original gains are fairly large (for relatively small opportunity cost initially), but gradually diminish (and opportunity cost of this additional production ultimately increases). Hence the law of diminishing returns or increasing opportunity cost. Similarly, as additional resources are added to healthcare, moving from bottom to top on the vertical axis, the original gains are fairly large, but again gradually diminish. In this way, the law of diminishing returns (law of increasing opportunity cost) produces the outward-bending shape of the production possibilities frontier.

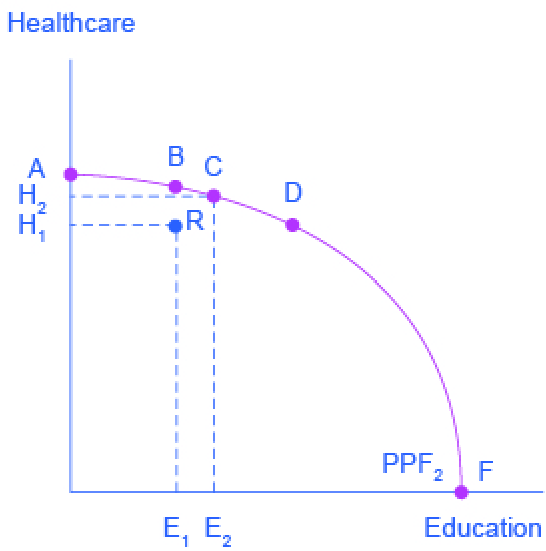
Productive Efficiency and Allocative Efficiency

The study of economics does not presume to tell a society what choice it should make along its production possibilities frontier. In a market-oriented economy with a democratic government, the choice will involve a mixture of decisions by individuals, firms, and government. However, economics can point out that some choices may be better than others. This observation is based on the concept of efficiency. In everyday usage, efficiency refers to lack of waste. An inefficient machine operates at high cost, while an efficient machine operates at lower cost, because it is not wasting energy or materials. An inefficient organization operates with long delays and high

costs, while an efficient organization meets schedules, is focused, and performs within budget.

The production possibilities curve/frontier is better at illustrating productive efficiency than allocative efficiency.

Productive and Allocative Efficiency



Productive efficiency means it is impossible to produce more of one good without decreasing the quantity that is produced of another good. Thus, all choices along a given PPF like B, C, and D display productive efficiency, but R does not. Allocative efficiency means that the particular mix of goods being produced—that is, the specific choice along the production possibilities frontier—represents the allocation that society most desires.

Productive efficiency means that, given the available inputs and technology, it is impossible to produce more of one good without decreasing the quantity that is produced of another good. All choices on the PPC/PPF in Figure 2, including A, B, C, D, and F, display productive efficiency. As a firm moves from any one of these choices to any other, either healthcare increases and education decreases or vice versa. However, any choice inside the production possibilities frontier is productively inefficient and wasteful because it is possible to produce more of one good, the other good, or some combination of both goods without decreasing output of neither good (hence no opportunity cost).

For example, point R is productively inefficient because it is possible at choice C to have more of both goods: education on the horizontal axis is higher at point C than point R (E_2 is greater than E_1), and healthcare on the vertical axis is also higher at point C than point R (H_2 is great than H_1).

Allocative efficiency means that the particular mix of goods a society produces represents the combination that society most desires. How to determine what a society desires can be a controversial question, and is usually discussed in political science, sociology, and philosophy classes as well as in economics. At its most basic, allocative efficiency means producers supply the quantity of each product that consumers demand. Only one of the productively efficient choices will be the allocatively efficient choice for society as a whole. Unless we are specifically told which particular combination of goods and services a society desires, the PPC/PPF model is really no good for illustrating allocative efficiency.

Why Society Must Choose

Every economy faces two situations in which it may be able to expand consumption of all goods. In the first case, a society may discover that it has been using its resources inefficiently, in which case by improving efficiency and producing more efficiently, it can move closer to a point on its PPC/PPF. In the second case, as resources grow over a period of years (e.g., more labor, more capital, more natural resources), the economy grows. As it does, the production possibilities frontier for a society will tend to shift outward and society will be able to afford more of all goods.

But improvements in productive efficiency take time to discover and implement, and economic growth happens only gradually. So, a society must choose between tradeoffs in the present. For government, this process often involves trying to identify where additional spending could do the most good and where reductions in spending would do the least harm. At the individual and firm level, the market economy coordinates a process in which firms seek to produce goods and services in the quantity, quality, and price that people want. But for both the government and the market economy in the short term, increases in production of one good typically mean offsetting decreases somewhere else in the economy.

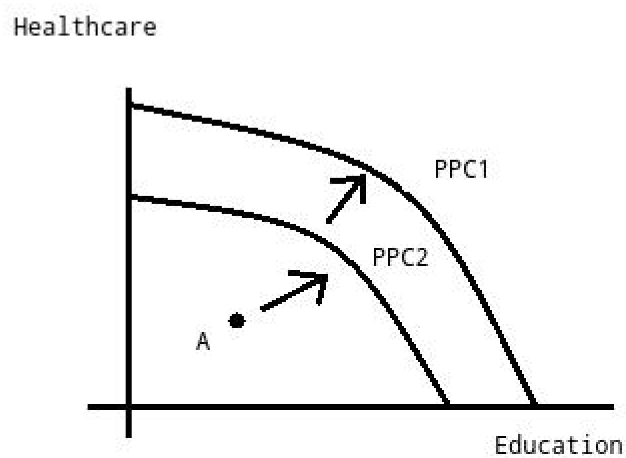


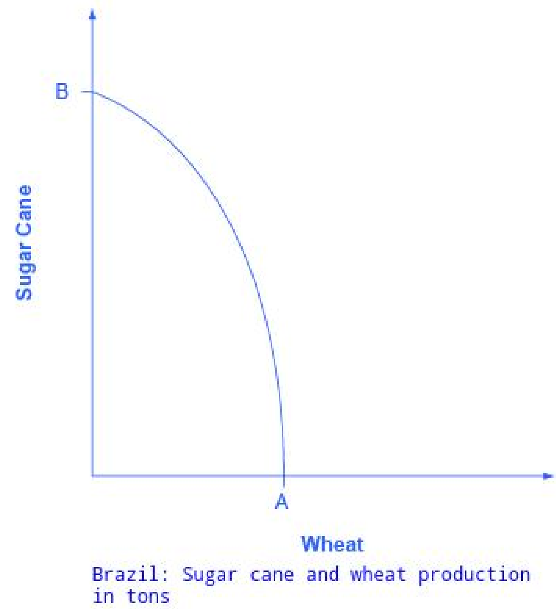
Figure 3. Increased productive efficiency will not result in a shift outward of the PPC/PPF (economic growth). It will result only in a move closer to the PPC e.g. from point A closer to the PPC. More resources on the other hand would result in economic growth e.g. a shift of the PPC/PPF from

PPC1 to PPC2. Declining growth would result in a shift inwards of the PPC from PPC2 to PPC1.

The PPF and Comparative Advantage

While every society must choose how much of each good it should produce, it does not need to produce every single good it consumes. Often how much of a good a country decides to produce depends on how expensive it is to produce it versus buying it from a different country. As we saw earlier, the curvature of a country's PPF gives us information about the tradeoff between devoting resources to producing one good versus another. In particular, its slope gives the opportunity cost of producing one more unit of the good in the x-axis in terms of the other good (in the y-axis). Countries tend to have different opportunity costs of producing a specific good, either because of different climates, geography, technology or skills.

Suppose two countries, South Africa and Brazil, need to decide how much they will produce of two crops: sugar cane and wheat. Due to its climatic conditions, Brazil can produce a lot of sugar cane per acre but not much wheat. Conversely, S.A can produce a lot of wheat per acre, but not much sugar cane. Clearly, Brazil has a lower opportunity cost of producing sugar cane (in terms of wheat) than South Africa. The reverse is also true; S.A. has a lower opportunity cost of producing wheat than Brazil. This can be illustrated by the PPFs of the two countries in Figure 4.



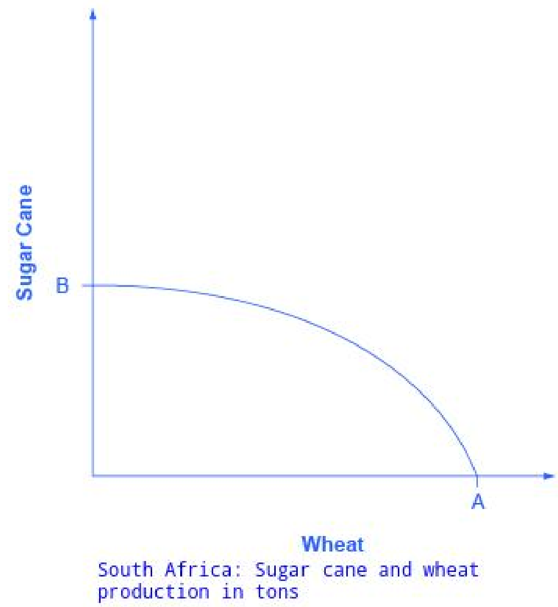


Figure 4: The South African PPF is flatter than the Brazil PPF. This means that South Africa has to give up less sugar cane to produce a lot of wheat implying that the opportunity cost of wheat (opportunity cost of producing for South Africa is less than it is for Brazil). Brazil, on the other hand has to give up a relatively small amount of wheat to produce a lot of sugar cane. So Brazil has a lower opportunity cost of producing sugar cane than South Africa has. So we say South Africa has comparative advantage in wheat and Brazil has comparative advantage in sugar cane.

When a country can produce a good at a lower opportunity cost than another country, we say that this country has a **comparative advantage** in that good. In our example, Brazil has a comparative advantage in sugar cane and South Africa has a comparative advantage in wheat. One can easily see this with a simple observation of the extreme production points in the PPFs of the two countries. If Brazil devoted all of its resources to producing wheat, it would be producing at point A. If however it had devoted all of its resources to producing sugar cane instead, it would be producing a much larger amount, at point B. By moving from point A to

point B Brazil would give up a relatively small quantity in wheat production to obtain a large production in sugar cane. The opposite is true for South Africa. If South Africa moved from point A to B and produced only sugar cane, this would result in a large opportunity cost in terms of foregone wheat production.

The slope of the PPC/PPF gives the opportunity cost of producing an additional unit of wheat. While the slope is not constant throughout the PPFs, it is clear that the PPF in Brazil is much steeper than in South Africa, and therefore the opportunity cost of wheat is generally higher in Brazil. In the chapter on International Trade you will learn that countries' differences in comparative advantage determine which goods they will choose to produce and trade. When countries engage in trade, they specialize in the production of the goods that they have comparative advantage in, and trade part of that production for goods they do not have comparative advantage in. With trade, goods are produced where the opportunity cost is lowest, so total production increases, benefiting both trading parties.

Key Concepts and Summary

A production possibilities curve/frontier defines the set of choices society faces for the combinations of goods and services it can produce given the resources available. The shape of the PPF is typically curved outward, rather than straight. Choices outside the PPF are unattainable and choices inside the PPF are wasteful. Over time, a growing economy will tend to shift the PPF outwards.

The law of diminishing returns holds that as additional amounts of resources are devoted to producing something, the marginal (extra or additional) increase in output will become smaller and smaller. All choices along a production possibilities frontier display productive efficiency; that is, it is impossible to use society's resources to produce more of one good without decreasing production of the other good. The specific choice along a production possibilities frontier that reflects the mix of goods society prefers is the choice with allocative efficiency. The curvature of the PPF is likely to differ by country, which results in different countries having comparative advantage in different goods. Total production can increase if

countries specialize in the goods they have comparative advantage in and trade some of their production for the remaining goods.

Self-Check Questions

Exercise:

Problem:

Return to the example in Figure 2. Suppose there is an improvement in medical technology that enables more healthcare to be provided with the same amount of resources. How would this affect the production possibilities curve and, in particular, how would it affect the opportunity cost of education?

Solution:

Because of the improvement in technology, the vertical intercept of the PPF would be at a higher level of healthcare. In other words, the PPF would rotate clockwise around the horizontal intercept. This would make the PPF steeper, corresponding to an increase in the opportunity cost of education, since resources devoted to education would now mean forgoing a greater quantity of healthcare.

Review Questions

Exercise:

Problem: What is comparative advantage?

Exercise:

Problem: What does a production possibilities frontier illustrate?

Exercise:

Problem:

Why is a production possibilities frontier typically drawn as a curve, rather than a straight line?

Exercise:

Problem:

Explain why societies cannot make a choice above their production possibilities frontier and should not make a choice below it.

Exercise:

Problem: What are diminishing marginal returns?

Exercise:

Problem: What is productive efficiency? Allocative efficiency?

Critical Thinking Questions

Exercise:

Problem:

During the Second World War, Germany’s factories were destroyed. It also suffered many human casualties, both soldiers and civilians. How did the war affect Germany’s production possibilities curve?

Exercise:

Problem:

It is clear that productive inefficiency is a waste since resources are being used in a way that produces less goods and services than a nation is capable of. Why is allocative inefficiency also wasteful?

Glossary

allocative efficiency
: when the mix of goods being produced represents the mix that society most desires

comparative advantage
: when a country can produce a good at a lower cost in terms of other goods; or, when a country has a lower opportunity cost of production

law of diminishing returns
: as additional increments of resources are added to producing a good or service, the marginal benefit from those additional increments will decline

production possibilities frontier (PPF)
: a diagram that shows the productively efficient combinations of two products that an economy can produce given the resources it has available.

productive efficiency
: when it is impossible to produce more of one good (or service) without decreasing the quantity produced of another good (or service)

Introduction to Demand and Supply
class="introduction"
Farmer’s Market

Organically grown (without using harmful pesticides and artificial fertilisers) vegetables and fruits that are produced and sold within a specific geographical region such as KwaZulu-Natal should, in theory, cost less than similar produce grown normally (non-organic) in another province (such as Gauteng) and sold locally in KZN because the transportation costs are less.

That is not, however, usually the case. (Credit: modification of work by Natalie Maynor/Flickr Creative Commons)

Note:
Why Can We Not Get Enough of Organic Produce?
Organic food is increasingly popular, not just in South Africa, but worldwide. Organically grown produce does not rely on harmful pesticides or artificial fertilizers and is less damaging to the environment. For example, it does not pollute the soil or water resources or kill the major pollinators of all plants, the bees. Many consumers now prefer organic

food as a healthier option. At one time, consumers had to go to specialty stores or farmer’s markets to find organic produce. Now it is available in some grocery stores. In short, organically grown food has become more in demand.

Ever wonder why organic food costs more than conventional food? Why, say, might an organic apple cost R30/kg, while normally grown apples might cost R20/kg? The same price relationship is true for just about every organic product on the market. If many organic foods are locally grown, would they not take less time to get to market and therefore be cheaper? What are the forces that keep those prices from coming down? Turns out those forces have a lot to do with this chapter’s topic: demand and supply.

Note:

Introduction to Demand and Supply

In this chapter, you will learn about:

- Demand, Supply, and Equilibrium in Markets for Goods and Services
- Shifts in Demand and Supply for Goods and Services
- Changes in Equilibrium Price and Quantity: The Four-Step Process
- Price Ceilings and Price Floors

A music lover pays pays thousands of Rands for Brenda Fassie's first music album of which only a few were produced. An art collector spends hundreds of thousands of Rands on a few small paintings produced by Irma Stern, a South African painter. People usually react to purchases like these in two ways: their jaw drops because they think these are high prices to pay for such goods or they think these are rare, desirable items and the amount paid seems right.

Note:

Visit this website to read a list of unusual items that have been purchased for their ties to celebrities. These examples represent an interesting facet of demand and supply.

When economists talk about prices, they are less interested in making judgements than in gaining a practical understanding of what determines prices and why prices change. Consider a price that concerns most of us: that of petrol. Why does the average price of petrol in South Africa change (mostly increase) every couple of months? To explain these price movements, economists focus on the determinants of what petrol buyers are willing to pay and what petrol sellers are willing to accept. Of course government taxes and the Rand-Dollar exchange rate can also influence fuel price movements. But we will get to that later. First things first...

This chapter introduces the economic model of demand and supply—one of the most powerful models in all of economics. The discussion here begins by examining how demand and supply determine the price and the quantity sold in markets for goods and services, and how changes in demand and supply lead to changes in prices and quantities.

Demand, Supply, and Equilibrium in Markets for Goods and Services

By the end of this section, you will be able to:

- Explain demand, quantity demanded, and the law of demand
- Identify a demand curve and a supply curve
- Explain supply, quantity supply, and the law of supply
- Explain equilibrium, equilibrium price, and equilibrium quantity

First let's first focus on what economists mean by demand, what they mean by supply, and then how demand and supply interact in a market.

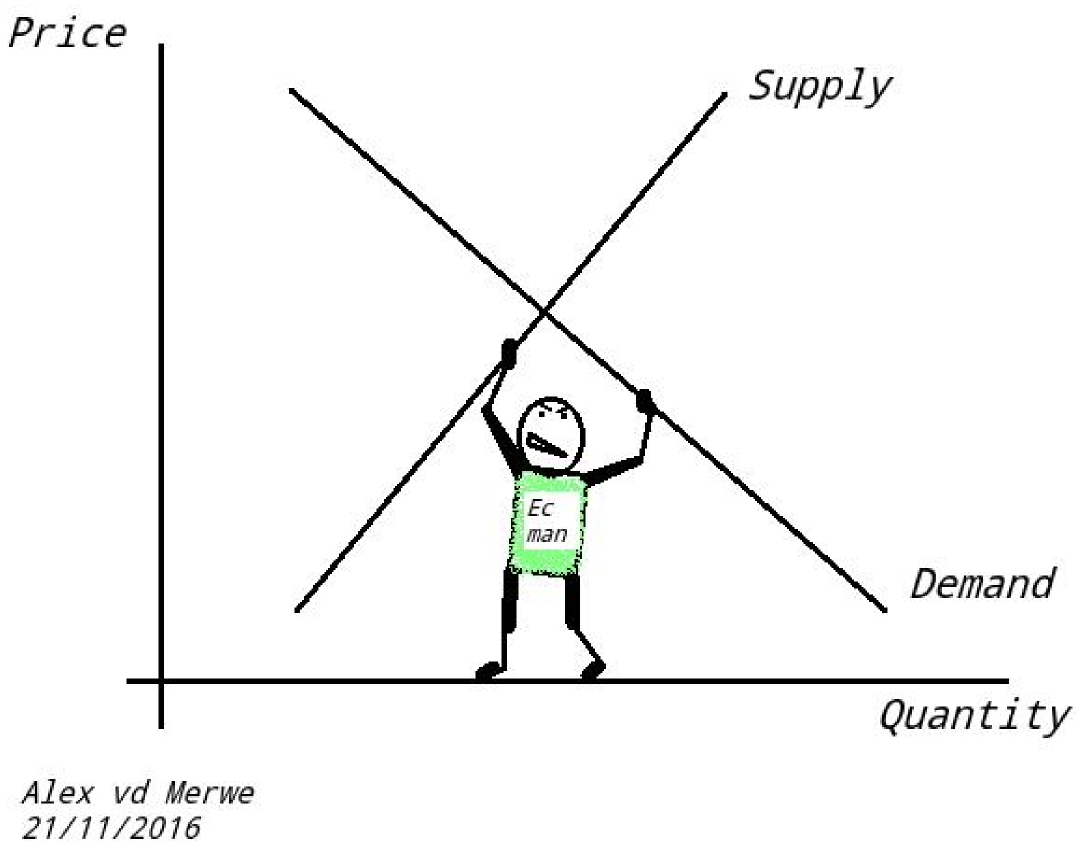


Figure 1. Econman building the famous supply and

demand model.

Demand for Goods and Services

Economists use the term **demand** to refer to the amount of some good or service consumers are willing and able to purchase at each price. Demand is based on needs and wants. However, demand is different from both needs and wants because it is based on ability to pay. If you cannot pay for it, you have no effective demand.

What a buyer pays for a unit of the specific good or service is called **price**. The total number of units purchased at that price is called the **quantity demanded**. A rise in price of a good or service almost always decreases the quantity demanded of that good or service. Conversely, a fall in price will increase the quantity demanded. So, for example, when the price of a litre of petrol increases people look for ways to reduce their consumption by combining several errands, sharing cars, using public transport and taking weekend or vacation trips closer to home. According to the law of demand, then, it is clear that quantity demanded depends on price (price is the independent variable and quantity demanded is the dependent variable). Economists call this inverse relationship between price and quantity demanded the **law of demand**. The law of demand assumes that all other variables that affect demand (to be explained in the next module) are held constant.

An example from the market for petrol can be shown in the form of a table or a graph. A table that shows the quantity demanded at each price, such as Table 1, is called a **demand schedule**. Price in this case is measured in Rands per litre of petrol. The quantity demanded is measured in millions of litres over some time period (for example, per day or per year) and over some geographic area (like a province or a country). A **demand curve** shows the relationship between price and quantity demanded on a graph like Figure 2, with quantity on the horizontal axis and the price per litre on the vertical axis. (Note that this is an exception to the normal rule in mathematics that the independent variable (x) goes on the horizontal axis and the dependent variable (y) goes on the vertical. Economics is not maths.)

The demand schedule shown by Table 1 and the demand curve shown by the graph in Figure 2 are two ways of describing the same relationship between price (P) and quantity demanded (Q).

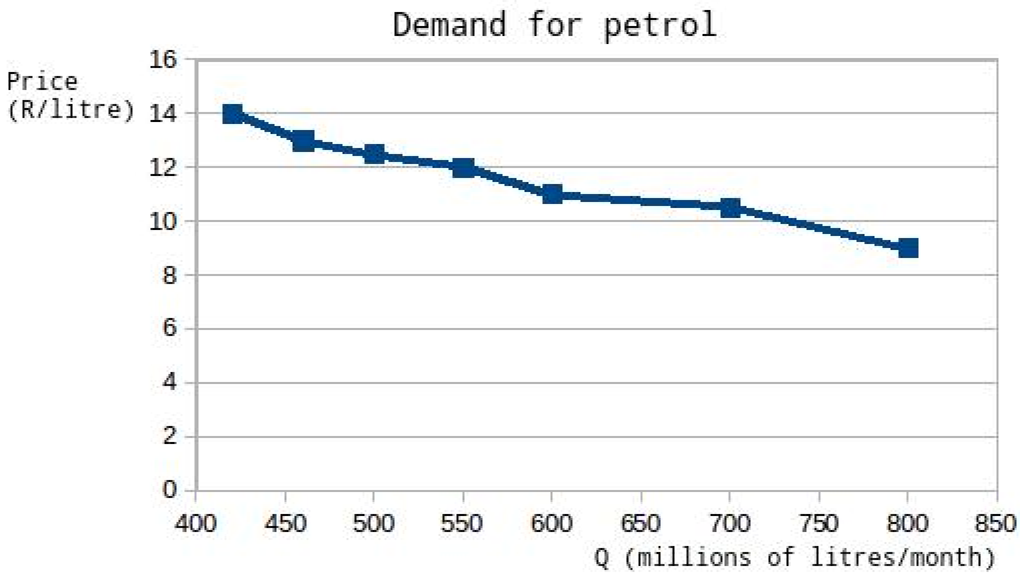


Figure 2: The demand schedule shows that as price (P) rises, quantity demanded (Q) decreases, and vice versa. These points are then graphed, and the line connecting them is the demand curve (D). The downward slope of the demand curve again illustrates the law of demand—the inverse relationship between prices and quantity demanded.

Price (per litre)	**Quantity Demanded (millions of litres/month)**
R9.00	800
R10.50	700
R11.00	600
R12.00	550
R12.50	500
R13.00	460
R14.00	420

Market Demand for Petrol

Demand curves will appear somewhat different for each product. They may appear relatively steep or flat, or they may be straight or curved. Nearly all demand curves share the fundamental similarity that they slope down from left to right. So demand curves embody the law of demand: As the price increases, the quantity demanded decreases, and conversely, as the price decreases, the quantity demanded increases. Demand curves may explain the consumption choices of individuals (individual demand) or of groups of people (market demand). Our focus in economic study is market demand although the law of demand obviously holds for individuals as well society as a whole (groups of people).

Confused about these different types of demand? Read the next Clear It Up feature.

Note:
Is demand the same as quantity demanded?

In economic terminology, demand is not the same as quantity demanded. When economists talk about demand, they mean the relationship between a range of prices and the quantities demanded at those prices, as illustrated by a demand curve or a demand schedule. When economists talk about quantity demanded, they mean only a certain point on the demand curve, or one quantity on the demand schedule. In short, demand refers to the curve and quantity demanded refers to the (specific) point on the curve.

Supply of Goods and Services

When economists talk about **supply**, they mean the amount of some good or service a producer is willing to supply at each price. Price is what the producer receives for selling one unit of a good or service. A rise in price almost always leads to an increase in the **quantity supplied** of that good or service, while a fall in price will decrease the quantity supplied. When the price of petrol rises, for example, it encourages profit-seeking firms to take several actions: expand exploration for oil reserves; drill for more oil; invest in more pipelines and oil tankers to bring the oil to plants where it can be refined into petrol; build new oil refineries; purchase additional pipelines and trucks to ship the petrol to fuel stations; and open more fuel stations or keep existing fuel stations open longer hours. Economists call this positive relationship between price and quantity supplied—that a higher price leads to a higher quantity supplied and a lower price leads to a lower quantity supplied—the **law of supply**. The law of supply assumes that all other variables that affect supply (to be explained in the next module) are held constant.

Still unsure about the different types of supply? See the following Clear It Up feature.

Note:
Is supply the same as quantity supplied?
In economic terminology, supply is not the same as quantity supplied. When economists refer to supply, they mean the relationship between a

range of prices and the quantities supplied at those prices, a relationship that can be illustrated with a supply curve or a supply schedule. When economists refer to quantity supplied, they mean only a certain point on the supply curve, or one quantity on the supply schedule. In short, supply refers to the curve and quantity supplied refers to the (specific) point on the curve.

Fig 3 illustrates the law of supply, again using the market for petrol as an example. Like demand, supply can be illustrated using a table or a graph. A **supply schedule** is a table, like Table 2, that shows the quantity supplied at a range of different prices. Again, price is measured in Rands per litre of petrol and quantity supplied is measured in millions of litres per month. A **supply curve** is a graphic illustration of the relationship between price, shown on the vertical axis, and quantity, shown on the horizontal axis. The supply schedule and the supply curve are just two different ways of showing the same information. Notice that the horizontal and vertical axes on the graph for the supply curve are the same as for the demand curve.

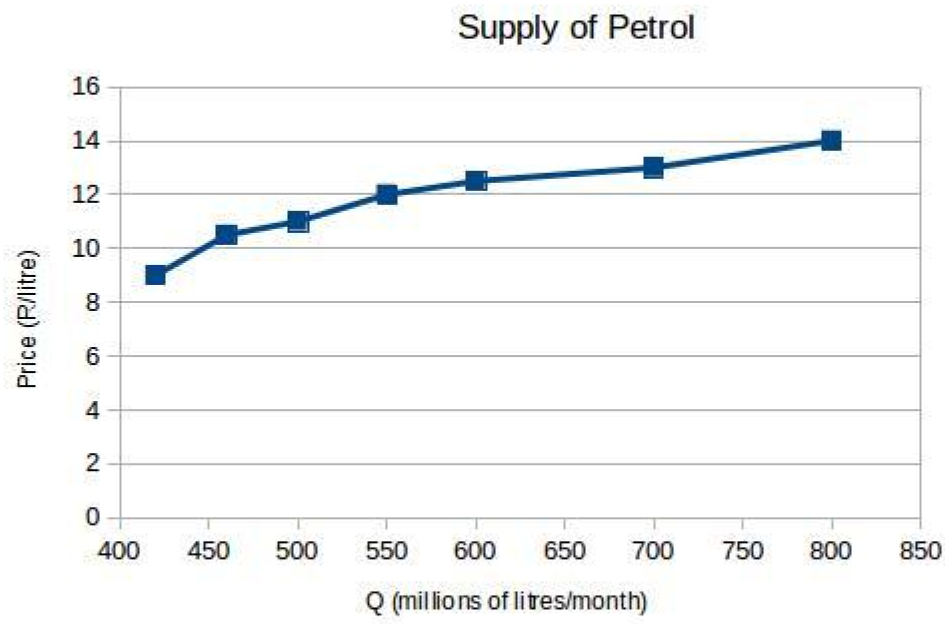


Figure 3: The supply schedule is the table that shows quantity supplied of petrol at each price. As price rises, quantity supplied also increases, and vice versa. The supply curve (S) is created by graphing the points from the supply schedule and then connecting them. The upward slope of the supply curve illustrates the law of supply—that a higher price leads to a higher quantity supplied, and vice versa.

Price (per litre)	Quantity Supplied (millions of litres/month)
R14.00	800
R13.00	700
R12.50	600
R12.00	550
R11.00	500
R10.50	460
R9.00	420

Market Supply of Petrol

The shape of supply curves will vary somewhat according to the product: steeper, flatter, straighter, or curved. Nearly all supply curves, however, share a basic similarity: they slope up from left to right and illustrate the

law of supply: as the price rises, say, from R9.00/litre to R12.00 per litre, the quantity supplied increases from 420 million litres/month to 550 million litres/month. Conversely, as the price falls, the quantity supplied decreases. The law of supply describes the behaviour of both individual firms or producers (individual supply) as well as market supply (groups of suppliers or the industry). Our focus is market supply.

Equilibrium—Where Demand and Supply Intersect

Because the graphs for demand and supply curves both have price on the vertical axis and quantity on the horizontal axis, the demand curve and supply curve for a particular good or service can appear on the same graph. Together, demand and supply determine the price and the quantity that will be bought and sold in a market.

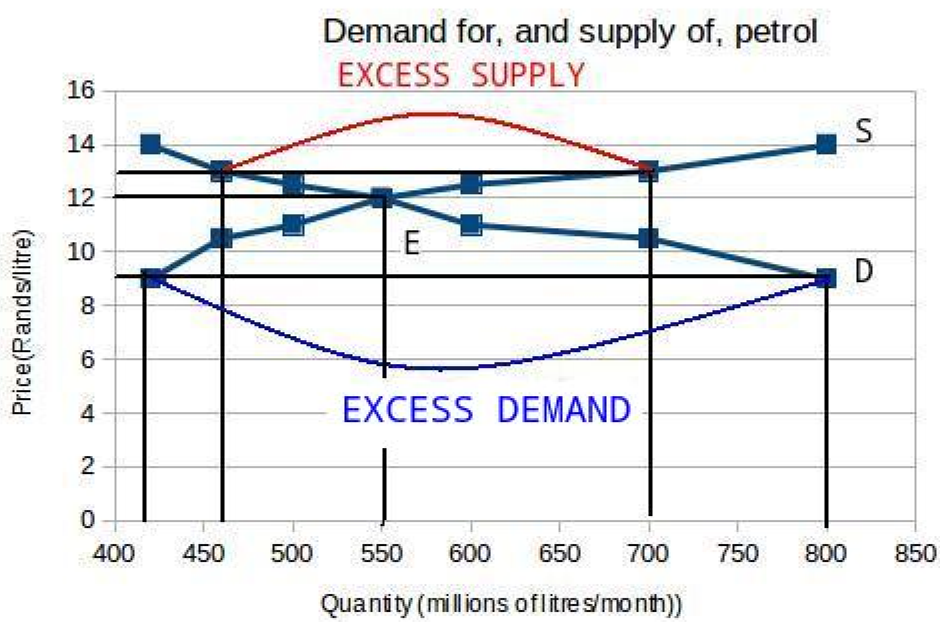


Figure 4: The demand curve (D) and the supply curve (S) intersect at the equilibrium point E, with a price of R12/litre and a quantity of 550

million litres/month. The equilibrium is the only price where quantity demanded is equal to quantity supplied. At a price above equilibrium like R13/litre, quantity supplied exceeds the quantity demanded, so there is excess supply. At a price below equilibrium such as R9/litre, quantity demanded exceeds quantity supplied, so there is excess demand.

Price (per litre)	**Quantity demanded (millions of litres/month)**	**Quantity supplied (millions of litres/month)**
R9.00	800	420
R10.50	700	460
R11.00	600	500
R12.00	550	550
R12.50	500	600
R13.00	460	700
R14.00	420	800

Price, Quantity Demanded, and Quantity Supplied

Remember this: When two lines on a diagram cross, this intersection usually means something. The point where the supply curve (S) and the demand curve (D) cross, designated by point E in Fig 4, is called the

equilibrium. The **equilibrium price** is the only price where the plans of consumers and the plans of producers agree—that is, where the amount of the product consumers want to buy (quantity demanded) is equal to the amount producers want to sell (quantity supplied). This common quantity is called the **equilibrium quantity**. At any other price, the quantity demanded does not equal the quantity supplied, so the market is not in equilibrium at that price.

In Fig 4, the equilibrium price is R12/litre of petrol and the equilibrium quantity is 550 million litres/month of petrol. If you had only the demand and supply schedules, and not the graph, you could find the equilibrium by looking for the price level on the tables where the quantity demanded and the quantity supplied are equal.

The word "equilibrium" means "balance." If a market is at its equilibrium price and quantity, then it has no reason to move away from that point. However, if a market is not at equilibrium, then economic pressures arise to move the market toward the equilibrium price and the equilibrium quantity.

Imagine, for example, that the price of a litre of petrol was above the equilibrium price—that is, instead of R12/litre, the price is R13/litre. At this higher price, the quantity demanded drops from 550 million litres/month to 460 million litres/month. This decline in quantity reflects how consumers react to the higher price by finding ways to use less petrol.

At this higher price of R13/litre, the quantity of petrol supplied rises from the 550 million litres/month to 700 million litres/month, as the higher price makes it more profitable for fuel producers and retailers to expand their output. Now, consider how quantity demanded and quantity supplied are related at this above-equilibrium price. Quantity demanded has fallen to 460 million litres/month, while quantity supplied has risen to 700 million litres/month. In fact, at any above-equilibrium price, the quantity supplied exceeds the quantity demanded. We call this an **excess supply** or a **surplus**.

With a surplus, petrol volumes accumulate at petrol stations, in tanker trucks, in pipelines, and at oil refineries. This accumulation or surplus puts pressure on petrol sellers. If a surplus remains unsold, those firms involved in making and selling petrol are not receiving enough cash to pay their

workers and to cover their expenses. In this situation, some producers and sellers will want to cut prices, because it is better to sell at a lower price than not to sell at all. Once some sellers start cutting prices, others will follow to avoid losing sales. These price reductions in turn will stimulate a higher quantity demanded. So, as long as the price is above the equilibrium level, temporary surpluses will build up which tend to push prices down towards equilibrium. This is the price adjustment mechanism which coordinates the intentions and actions of buyers and sellers in all markets.

Now suppose that the price is below its equilibrium level at R9/litre. At this lower price, the quantity demanded increases from 550 million litres/month to 800 million litres/month as drivers take longer trips, spend more minutes warming up the car in the driveway in wintertime, stop sharing rides to work, and buy larger cars that consume more petrol per kilometre. However, the below-equilibrium price reduces petrol producers' incentives to produce and sell petrol, and the quantity supplied falls from 550 million litres/month to 420 million litres/month.

When the price is below equilibrium, there is **excess demand**, or a **shortage**—that is, at the given price the quantity demanded, which has been stimulated by the lower price, now exceeds the quantity supplied, which had been depressed by the lower price. In this situation, eager petrol buyers rush to the petrol stations, only to find many stations running short of fuel. Oil companies and petrol stations recognize that they have an opportunity to make higher profits by selling what petrol they do have at a higher price. As a result, the price rises toward the equilibrium level. Read Demand, Supply, and Efficiency for more discussion on the importance of the demand and supply model.

Again, we see the price adjustment mechanism at work. At prices below equilibrium temporary shortages develop. These temporary shortages cause price to increase towards equilibrium. As price increases so the quantity demanded decreases (law of demand) and quantity supplied increases (law of supply) until the shortage is eliminated. When there is no longer any shortage, there is no further pressure on price to adjust. Price comes to rest when quantity demanded exactly matches quantity supplied (equilibrium).

At equilibrium there is no temporary shortage or surplus to force price to adjust in an effort to rebalance the quantities traded.

Key Concepts and Summary

A demand schedule is a table that shows the quantity demanded at different prices either by an individual (individual demand) or in a market (market demand). A market demand curve shows the relationship between quantity demanded and price in a given market on a graph. The law of demand states that a higher price typically leads to a lower quantity demanded.

A supply schedule is a table that shows the quantity supplied at different prices either by one supplier (individual supply) or by all suppliers (market supply). A market supply curve shows the relationship between quantity supplied by all firms and price on a graph. The law of supply states that a higher price typically leads to a higher quantity supplied.

The equilibrium price and equilibrium quantity occur where the supply and demand curves cross. The equilibrium occurs where the quantity demanded is equal to the quantity supplied. If the price is below the equilibrium level, then the quantity demanded will exceed the quantity supplied. Excess demand or a shortage will exist. If the price is above the equilibrium level, then the quantity supplied will exceed the quantity demanded. Excess supply or a surplus will exist. In the case of excess demand, temporary shortages will push the price up toward the equilibrium level. In the case of excess supply, temporary surpluses will push the price down toward the equilibrium level. This is known as the **price adjustment mechanism**.

Self-Check Question

Exercise:

Problem:

Review Fig 4. Suppose the price of petrol is R12.50 per litre. Is the quantity demanded higher or lower than at the equilibrium price of R12 per litre? And what about the quantity supplied? Is there a shortage or a surplus in the market? If so, of how much?

Solution:

Since R12.50 per litre is above the equilibrium price, the quantity demanded would be lower at 500 million litres per month and the quantity supplied would be higher at 600 litres per month. (These results are due to the laws of demand and supply, respectively.) The outcome of lower Quantity demanded (Qd) and higher Quantity supplied (Qs) would be a surplus in the petrol market of 600 – 500 = 100 million litres per month.

Review Questions

Exercise:

Problem: What determines the level of prices in a market?

Exercise:

Problem:

What does a downward-sloping demand curve mean about how buyers in a market will react to a higher price?

Exercise:

Problem:

Will demand curves have the same exact shape in all markets? If not, how will they differ?

Exercise:

Problem:

Will supply curves have the same shape in all markets? If not, how will they differ?

Exercise:

Problem:

What is the relationship between quantity demanded and quantity supplied at equilibrium? What is the relationship when there is a shortage? What is the relationship when there is a surplus?

Exercise:

Problem:

How can you locate the equilibrium point on a demand and supply graph?

Exercise:

Problem:

If the price is above the equilibrium level, would you predict a surplus or a shortage? If the price is below the equilibrium level, would you predict a surplus or a shortage? Why?

Exercise:

Problem:

When the price is above the equilibrium, explain how market forces move the market price to equilibrium. Do the same when the price is below the equilibrium.

Exercise:

Problem:

What is the difference between the demand and the quantity demanded of a product, say milk? Explain in words and show the difference on a graph with a demand curve for milk.

Exercise:

Problem:

What is the difference between the supply and the quantity supplied of a product, say milk? Explain in words and show the difference on a graph with the supply curve for milk.

Critical Thinking Questions

Exercise:

Problem:

Review Fig 4. Suppose the government decided that, since petrol is a necessity, its price should be legally capped and restricted to a maximum price of R10.50 per litre. What do you anticipate would be the outcome in the petrol market?

Exercise:

Problem:

Explain why the following statement is false: “In the goods market, no buyer would be willing to pay more than the equilibrium price.”

Exercise:

Problem:

Explain why the following statement is false: “In the goods market, no seller would be willing to sell for less than the equilibrium price.”

Problems

Exercise:

Problem:

Review Fig 4 again. Suppose the price of petrol is R9 per litre. Will the quantity demanded be lower or higher than at the equilibrium price of R12 per litre? Will the quantity supplied be lower or higher? Is there a shortage or a surplus in the market? If so, of how much?

Glossary

demand curve
: a graphic representation of the relationship between price and quantity demanded of a certain good or service, with quantity on the horizontal axis and the price on the vertical axis

demand schedule
: a table that shows a range of prices for a certain good or service and the quantity demanded at each price

demand
: the relationship between price and the quantity demanded of a certain good or service

equilibrium price
: the price where quantity demanded is equal to quantity supplied

equilibrium quantity
: the quantity at which quantity demanded and quantity supplied are equal for a certain price level

equilibrium
: the situation where quantity demanded is equal to the quantity supplied; the combination of price and quantity where there is no economic pressure from surpluses or shortages that would cause price or quantity to change

excess demand

at the existing price, the quantity demanded exceeds the quantity supplied; also called a shortage

excess supply
: at the existing price, quantity supplied exceeds the quantity demanded; also called a surplus

law of demand
: the common relationship that a higher price leads to a lower quantity demanded of a certain good or service and a lower price leads to a higher quantity demanded, while all other variables are held constant

law of supply
: the common relationship that a higher price leads to a greater quantity supplied and a lower price leads to a lower quantity supplied, while all other variables are held constant

price
: what a buyer pays for a unit of the specific good or service

quantity demanded
: the total number of units of a good or service consumers are willing to purchase at a given price

quantity supplied
: the total number of units of a good or service producers are willing to sell at a given price

shortage
: at the existing price, the quantity demanded exceeds the quantity supplied; also called excess demand

supply curve
: a line that shows the relationship between price and quantity supplied on a graph, with quantity supplied on the horizontal axis and price on the vertical axis

supply schedule

a table that shows a range of prices for a good or service and the quantity supplied at each price

supply
: the relationship between price and the quantity supplied of a certain good or service

surplus
: at the existing price, quantity supplied exceeds the quantity demanded; also called excess supply

Shifts in Demand and Supply for Goods and Services

By the end of this section, you will be able to:

- Identify factors that affect demand
- Graph demand curves and demand shifts
- Identify factors that affect supply
- Graph supply curves and supply shifts

The previous module explored how price affects the quantity demanded and the quantity supplied. The result was the demand curve and the supply curve. Price, however, is not the only thing that influences demand. Nor is it the only thing that influences supply. For example, how is demand for vegetarian food affected if, say, health concerns cause more consumers to avoid eating meat? Or how is the supply of diamonds affected if diamond producers discover several new diamond mines? What are the major factors, in addition to the price, that influence demand or supply?

Note:
Visit this website to read a brief note on how marketing strategies can influence supply and demand of products.

What Factors Affect Demand?

We defined demand as the amount of some product a consumer is willing and able to purchase at each price. That suggests at least two factors in addition to price that affect demand. Willingness to purchase suggests a

desire, based on what economists call tastes and preferences. If you neither need nor want something, you will not buy it. Ability to purchase suggests that income is important. Professors are usually able to afford better housing and transportation than students, because they have more income. Prices of related goods can affect demand also. If you need a new car, the price of a Honda may affect your demand for a Ford. Finally, the size or composition of the population can affect demand. The more children a family has, the greater their demand for clothing. The more driving-age children a family has, the greater their demand for car insurance, and the less for nappies and baby formula.

These factors matter both for demand by an individual and demand by the market as a whole. Exactly how do these various factors affect demand, and how do we show the effects graphically? To answer those questions, we need the *ceteris paribus* assumption.

The *Ceteris Paribus* Assumption

A demand curve or a supply curve is a relationship between two, and only two, variables: quantity on the horizontal axis and price on the vertical axis. The assumption behind a demand curve or a supply curve is that no relevant economic factors, other than the product's price, are changing. Economists call this assumption **ceteris paribus**, a Latin phrase meaning "other things being equal." Any given demand or supply curve is based on the *ceteris paribus* assumption that all else is held equal. A demand curve or a supply curve is a relationship between two, and only two, variables when all other variables are kept constant. If all else is not held equal, then the laws of supply and demand will not necessarily hold, as the following Clear It Up feature shows.

Note:
When does *ceteris paribus* apply?
Ceteris paribus is typically applied when we look at how changes in price affect demand or supply, but *ceteris paribus* can be applied more generally. In the real world, demand and supply depend on more factors than just

price. For example, a consumer's demand depends on income and a producer's supply depends on the cost of producing the product. How can we analyze the effect on demand or supply if multiple factors are changing at the same time—say price rises and income falls? The answer is that we examine the changes one at a time, assuming the other factors are held constant.

For example, we can say that an increase in the price reduces the amount consumers will buy (assuming income, and anything else that affects demand, is unchanged). Additionally, a decrease in income reduces the amount consumers can afford to buy (assuming price, and anything else that affects demand, is unchanged). This is what the *ceteris paribus* assumption really means. In this particular case, after we analyze each factor separately, we can combine the results. The amount consumers buy falls for two reasons: first because of the higher price and second because of the lower income.

How Does Income Affect Demand?

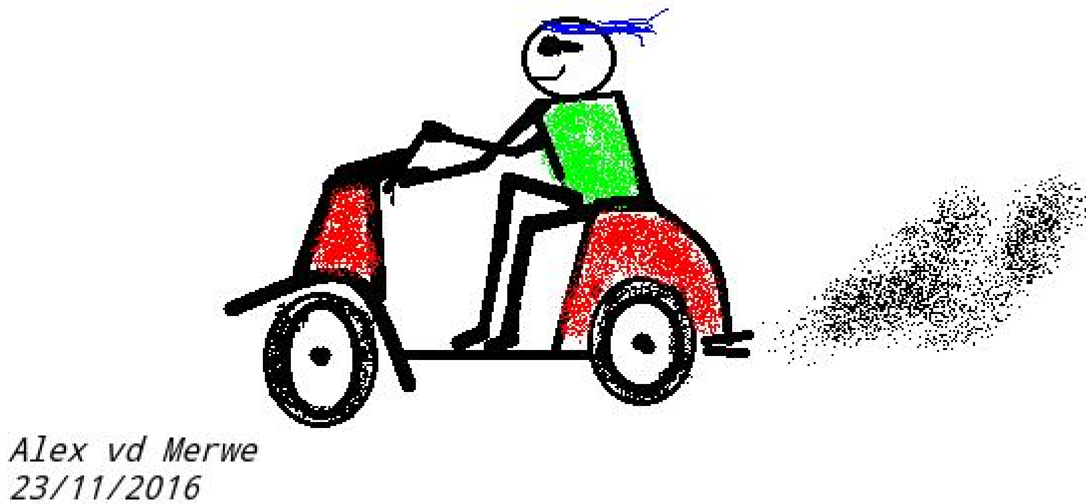


Figure 1

Let's use income as an example of how factors other than price affect demand. Figure 2 shows the initial demand for scooters as D_0. At point Q, for example, if the price is R20,000 per scooter, the quantity of scooters demanded is 18 million. D_0 also shows how the quantity of scooters demanded would change as a result of a higher or lower price. For example, if the price of a scooter rose to R22,000, the quantity demanded would decrease to 17 million, at point R.

The original demand curve D_0, like every demand curve, is based on the *ceteris paribus* assumption that no other economically relevant factors change. Now imagine that the economy expands in a way that raises the incomes of many people, making scooters more affordable. How will this affect demand? How can we show this graphically?

Return to Figure 2. The price of scooters is still R20,000, but with higher incomes, the quantity demanded has now increased to 20 million scooters, shown at point S. As a result of the higher income levels, the demand curve

shifts to the right to the new demand curve D_1, indicating an increase in demand. Table 1 below - *Prices and Shifts in Demand: A Scooters example* shows clearly that this increased demand would occur at every price, not just the original one.

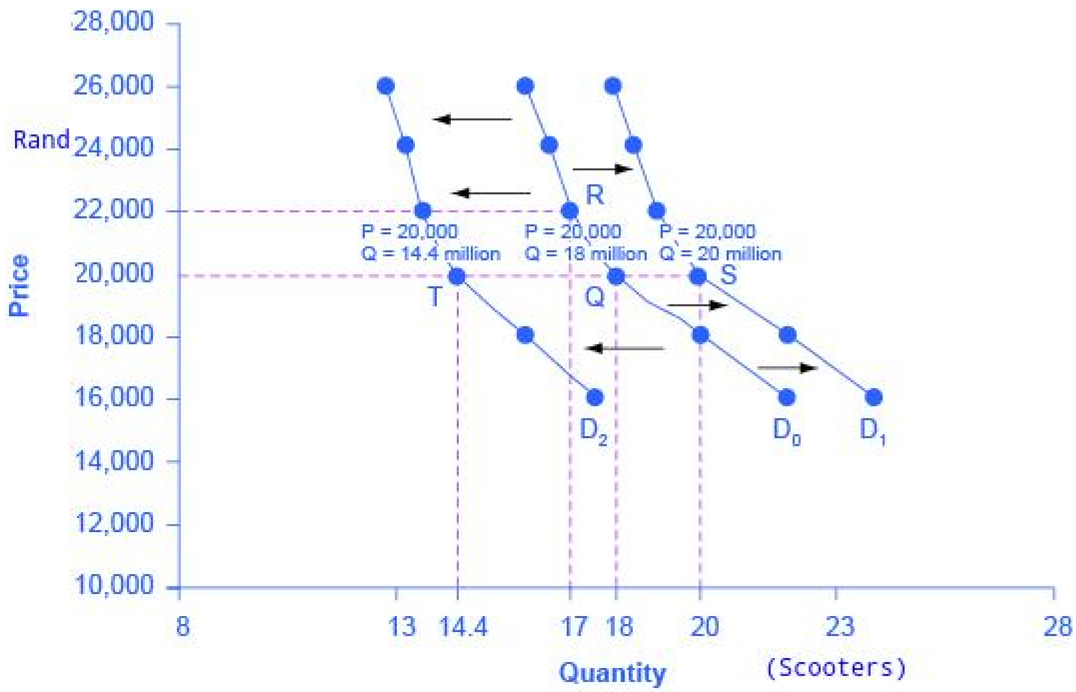


Figure 2: Increased demand means that at every given price, the quantity demanded is higher, so that the demand curve shifts to the right from D0 to D1. Decreased demand means that at every given price, the quantity demanded is lower, so that the demand curve shifts to the left from D0 to D2.

Prices and shifts in demand: A Scooter Example

Price (R/scooter)	Decrease to D2	Original demand D0	Increase to D1
16 000	17.6m	22m	24m
18 000	16m	20m	22m
20 000	14.4m	18m	20m
22 000	13.6m	17m	19m
24 000	13.2m	16.5m	18.5m
26 000	12.8m	16m	18m

Table 1. Prices and shifts in demand: a scooter example.

Now, imagine that the economy slows down so that many people lose their jobs or work fewer hours, reducing their incomes. In this case, the decrease in income would lead to a lower quantity of scooters demanded at every given price, and the original demand curve D_0 would shift left to D_2. The shift from D_0 to D_2 represents such a decrease in demand: At any given price level, the quantity demanded is now lower. In this example, a price of R20,000 means 18 million scooters sold along the original demand curve, but only 14.4 million scooters sold after demand fell.

When a demand curve shifts, it does not mean that the quantity demanded by every individual buyer changes by the same amount. In this example, not everyone would have higher or lower income and not everyone would buy or not buy an additional car. Instead, a shift in a demand curve captures a pattern for the market as a whole.

In the previous section, we argued that higher income causes greater demand at every price. This is true for most goods and services. For some—luxury cars, vacations in Europe, and fine jewelry—the effect of a rise in income can be especially large. A product whose demand rises when income rises, and vice versa, is called a **normal good**. A few exceptions to this pattern do exist. As incomes rise, many people will buy fewer generic (no name) brand groceries and more name brand groceries. They are less likely to buy used scooters or cars and more likely to buy new scooters or cars. They will be less likely to rent a flat or apartment and more likely to own a home, and so on. A product whose demand falls when income rises, and vice versa, is called an **inferior good**. In other words, when income increases, the demand curve shifts to the left.

Other Factors That Shift Demand Curves

Income is not the only factor that causes a shift in demand. Other things that change demand include tastes and preferences, the composition or size of the population, the prices of related goods, and even expectations. A change in any one of the underlying factors that determine what quantity people are willing to buy at a given price will cause a shift in demand. Graphically, the new demand curve lies either to the right (an increase) or to the left (a decrease) of the original demand curve. Let's look at these factors.

Changing Tastes or Preferences

The consumption of poultry meat in South Africa increased by almost 80%, from 21.5 kg per person per year in 2000 to 38.5 kg per person per year in 2014. As chicken is relatively inexpensive and generally available in most places, it has grown to be the most important protein source in the diet of the majority of South Africans (Burgin: 2015). This development has been attributed to steady economic growth and an increase in average income levels of South Africans since 2000. Developments like these change the quantity of a good demanded at every price: that is, they shift the demand curve for that good (in this case chicken meat) rightward.

Changes in the Composition of the Population

According to the World Bank (2015), South Africa is experiencing a significant demographic shift in which the share of its working-age population between 15 and 64 years has expanded substantially and will continue to grow for another five decades. Since 1994, the working age population has grown by 11 million and comprises 65% of South Africa's population of 54.9 million in 2015. It is expected that the working age population will grow by another 9 million in the next 50-years. More people in employment should result in more production and more income and more spending (demand). The World Bank report argues that this expansion of South Africa's workforce presents the country with a "demographic window of opportunity" for increased economic growth and better living standards.

The demand for a product can also be affected by changes in the prices of related goods such as substitutes or complements. A **substitute** is a good or

service that can be used in place of another good or service. As electronic books, like this one, become more available, you would expect to see a decrease in demand for traditional printed books. A lower price for a substitute decreases demand for the other product. For example, in recent years as the price of tablet computers has fallen, the quantity demanded has increased (because of the law of demand). Since people are purchasing tablets, there has been a decrease in demand for laptops, which can be shown graphically as a leftward shift in the demand curve for laptops. A higher price for a substitute good has the reverse effect.

Other goods are **complements** for each other, meaning that the goods are often used together, because consumption of one good tends to increase consumption of the other. Examples include: breakfast cereal and milk; notebooks and pens or pencils; golf balls and golf clubs; petrol and cars; and bread and butter/margarine. If the price of golf clubs rises, the quantity of golf clubs demanded falls (because of the law of demand) and demand for a complement good like golf balls will decrease too. Similarly, a higher price for diesel fuel would shift the demand curve for a complement good like diesel vehicles to the left. A lower price for a complement good has the reverse effect.

Changes in Expectations about Future Prices or Other Factors that Affect Demand

While it is clear that the price of a good affects the quantity demanded, it is also true that expectations about the future price (or expectations about tastes and preferences, income, and so on) can affect demand. For example, if people hear that Eskom will be reintroducing load shedding for the next two years, they may rush to the store to buy flashlight batteries or generators. If people learn that the price of a good like coffee is likely to rise in the future, they may head for the store to stock up on coffee now. These changes in demand are shown as shifts in the curve. Therefore, a **shift in demand** happens when a change in some economic factor (other than price of the good itself) causes a different quantity to be demanded at every price. The following Work It Out feature shows how this happens.

Note:

Shift in Demand

A shift in demand means that at any price (and at every price), the quantity demanded will be different than it was before. Here is an example showing a shift in demand for a normal good like pizza due to an income increase.

Step 1. Draw the graph of a demand curve for a normal good like pizza. Pick a price (like P_0). Identify the corresponding Q_0. An example is shown in Figure 3.

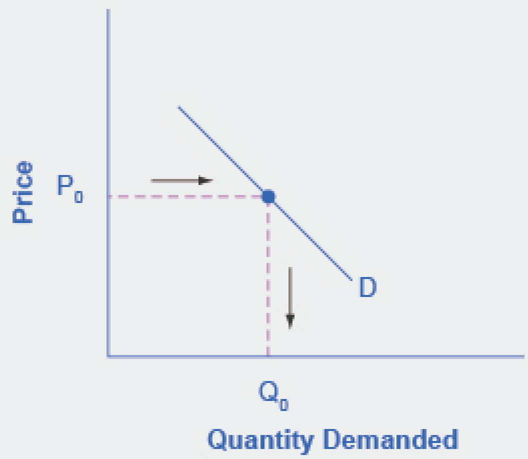


Figure 3: The demand curve can be used to identify how much consumers would buy at any given price.

Step 2. Suppose income increases. As a result of the change, are consumers going to buy more or less pizza? The answer is more. Draw a dotted horizontal line from the chosen price, through the original quantity demanded, to the new point with the new Q_1. Draw a dotted vertical line down to the horizontal axis and label the new Q_1. An example is provided in Figure 4.

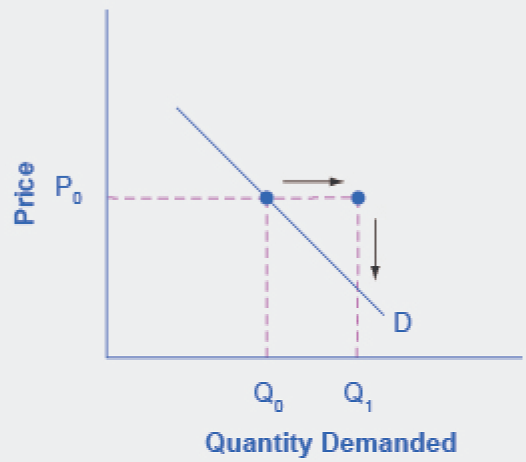


Figure 4: With an increase in income, consumers will purchase larger quantities, pushing demand to the right.

Step 3. Now, shift the curve through the new point. You will see that an increase in income causes a rightward shift in the demand curve, so that at any price the quantities demanded will be higher, as shown in Figure 5.

Figure 5: With an increase in income, consumers will purchase larger quantities, pushing demand to

```
the right, and causing the demand curve to shift
right.
```

Summing Up: Factors That Change Demand

Six factors that can shift demand curves are summarized in Figure 6. The direction of the arrows indicates whether the demand curve shifts represent an increase in demand or a decrease in demand. Notice that a change in the price of the good or service itself is not listed among the factors that can shift a demand curve. A change in the price of a good or service itself causes a movement along a specific demand curve, and it typically leads to some change in the quantity demanded, but it does not shift the demand curve.

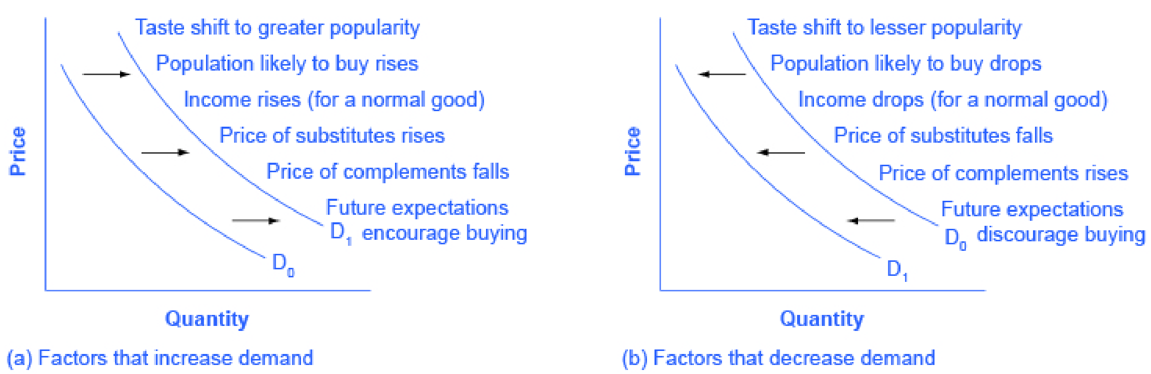


Figure 6: (a) A list of factors that can cause an increase in demand from D0 to D1. (b) The same factors, if their direction is reversed, can cause a decrease in demand from D0 to D1.

When a demand curve shifts, it will then intersect with a given supply curve at a different equilibrium price and quantity. We are, however, getting ahead of our story. Before discussing how changes in demand can affect

equilibrium price and quantity, we first need to discuss shifts in supply curves.

How Production Costs Affect Supply

A supply curve shows how quantity supplied will change as the price rises and falls, assuming *ceteris paribus* that no other economically relevant factors are changing. If other factors relevant to supply do change, then the entire supply curve will shift. Just as a shift in demand is represented by a change in the quantity demanded at every price, a **shift in supply** means a change in the quantity supplied at every price.

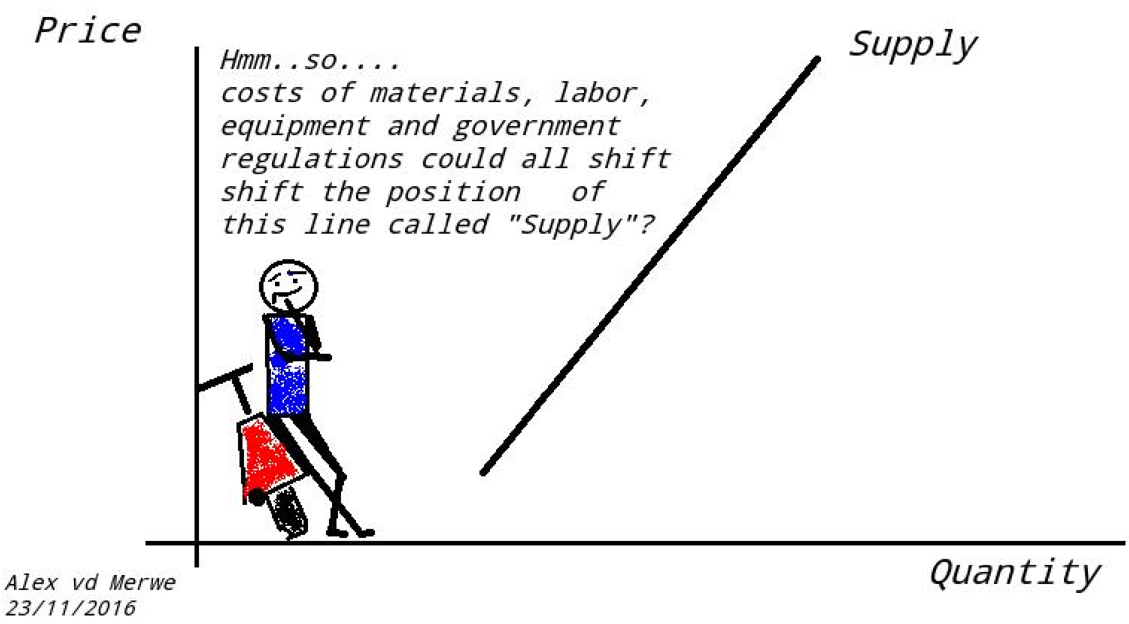


Figure 7.

In thinking about the factors that affect supply, remember what motivates firms: profits, which are the difference between revenues and costs. Goods and services are produced using combinations of labor, materials, and machinery, or what we call **inputs** or **resources, or factors of production**.

If a firm faces lower costs of production, while the prices for the good or service the firm produces remain unchanged, a firm’s profits go up. When a firm’s profits increase, it is more motivated to produce output, since the more it produces the more profit it will earn. So, when costs of production fall, a firm will tend to supply a larger quantity at any given price for its output. This can be shown by the supply curve shifting to the right.

Take, for example, a courier service that delivers packages around a city. The company or firm may find that buying fuel is one of its main costs. If the price of fuel falls, then the company will find it can deliver parcels and packages more cheaply than before. Since lower costs correspond to higher profits, the courier firm or company may now supply more of its services at any given price. For example, given the lower fuel prices, the company can now serve a greater area, and increase its supply.

Conversely, if a firm faces higher costs of production, then it will earn lower profits at any given selling price for its products. As a result, a higher cost of production typically causes a firm to supply a smaller quantity at any given price. In this case, the supply curve shifts to the left.

Consider the supply of scooters, shown by curve S_0 in Fig 8. Point J indicates that if the price is R20,000, the quantity supplied will be 18 million scooters. If the price rises to R22,000 per scooter, *ceteris paribus*, the quantity supplied will rise to 20 million scooters, as point K on the S_0 curve shows. The same information can be shown in table form, as in Table 2 below called *Prices and Shifts in Supply: A Scooter Example*

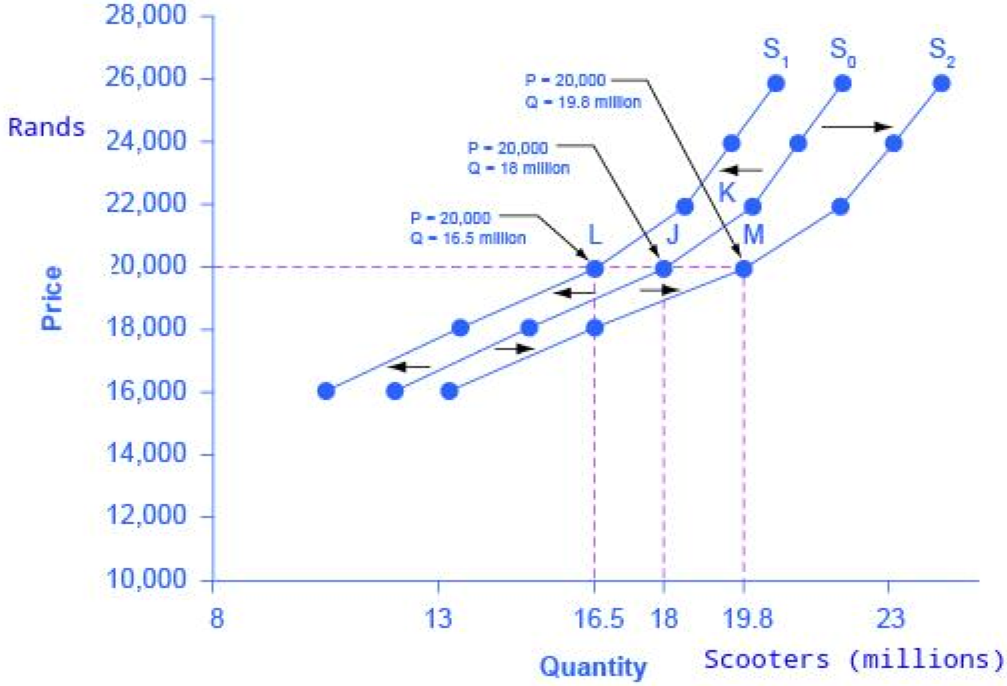


Figure 8: Decreased supply means that at every given price, the quantity supplied is lower, so that the supply curve shifts to the left, from S0 to S1. Increased supply means that at every given price, the quantity supplied is higher, so that the supply curve shifts to the right, from S0 to S2.

Prices and shifts in supply: A Scooter Example

Price (R/scooter)	Decrease to S1	Original Supply S0	Increase to S2
16 000	10.5m	12m	13.2m
18 000	13.5m	15m	16.5m
20 000	1.5m	18m	19.8m
22 000	18.5m	20m	22m
24 000	19.5m	21m	23.1m
26 000	20.5m	22m	24.2m

Table 2. Prices and shifts in supply: a scooter example.

Now, imagine that the price of steel, an important input in manufacturing scooters, rises, so that producing a scooter has become more expensive. At any given price for selling scooters, manufacturers will react by supplying a lower quantity. This can be shown graphically as a leftward shift of supply, from S_0 to S_1, which indicates that at any given price, the quantity supplied decreases. In this example, at a price of R20,000, the quantity supplied decreases from 18 million on the original supply curve (S_0) to 16.5 million on the supply curve S_1, which is labeled as point L.

Conversely, if the price of steel decreases, producing a scooter becomes less expensive. At any given price for selling scooters, manufacturers can now expect to earn higher profits, so they will supply a higher quantity. The shift of supply to the right, from S_0 to S_2, means that at all prices, the quantity supplied has increased. In this example, at a price of R20,000, the quantity supplied increases from 18 million on the original supply curve (S_0) to 19.8 million on the supply curve S_2, which is labeled M.

Other Factors That Affect Supply

In the example above, we saw that changes in the prices of inputs in the production process will affect the cost of production and thus the supply. Several other things affect the cost of production, too, such as changes in weather or other natural conditions, new technologies for production, and some government policies.

The cost of production for many agricultural products including food is affected by changes in natural conditions. For example, South Africa experienced its worst drought in 100 years in 2015/2016 (Vollgraaf and Mbatha: 2016). This has caused not only agricultural output to fall drastically but manufacturing is also negatively affected. A drought decreases the supply of agricultural products, which means that at any given price, a lower quantity (or poorer quality) will be supplied; conversely, good weather would shift the supply curve to the right.

When a firm discovers a new technology or method that allows it to produce at a lower cost, the supply curve will shift to the right. For instance, in the 1960s a major scientific effort known as the Green Revolution

focused on breeding improved seeds for basic crops like wheat and rice. By the early 1990s, more than two-thirds of the wheat and rice in low-income countries around the world was grown with these new improved seeds—and the harvest was twice as high per hectare. A technological improvement that reduces costs of production will shift supply to the right, so that a greater quantity will be produced at any given price.

Government policies can affect the cost of production and the supply curve through taxes, regulations, and subsidies. For example, the South African government each year imposes special taxes or levies on products such as petrol, alcohol and cigarettes. Taxes are treated as costs by businesses. Higher costs decrease supply for the reasons discussed above. Other examples of government policy that can affect cost are the wide range of government regulations that require firms to spend money to provide a cleaner environment or a safer and fairer workplace; complying with regulations (whether for safety, environmental or labor) increases firms' costs of business.

A government subsidy, on the other hand, is the opposite of a tax. A subsidy occurs when the government pays a firm directly or reduces the firm’s taxes if the firm carries out certain actions. South Africa's public universities, for example, receive a limited subsidy from the state for the purpose of reducing the cost of higher education for students. From the firm’s perspective, taxes or regulations are an additional cost of production that shifts supply to the left, leading the firm to produce a lower quantity at every given price. Government subsidies, on the other hand, reduce the cost of production and increase supply at every given price, shifting supply to the right. The following Work It Out feature shows how this shift happens.

Note:
Shift in Supply
We know that a supply curve shows the minimum price a firm will accept to produce a given quantity of output. What happens to the supply curve when the cost of production goes up? Here is an example of a shift in supply of pizzas due to a production cost increase.

Step 1. Draw a graph of a supply curve for pizza. Pick a quantity (like Q_0). If you draw a vertical line up from Q_0 to the supply curve, you will see the price the firm chooses. An example is shown in Figure 9.

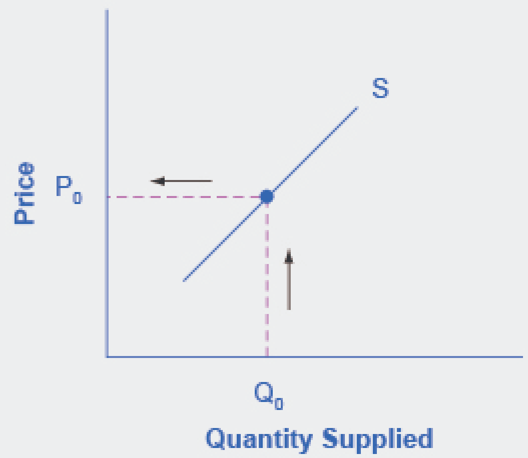


Figure 9: The supply curve can be used to show the minimum price a firm will accept to produce a given quantity of output.

Step 2. Why did the firm choose that price and not some other? One way to think about this is that the price is composed of two parts. The first part is the average cost of production, in this case, the cost of the pizza ingredients (flour, tomato sauce, cheese, herbs, and so on), the cost of heating the pizza oven, the rent on the shop, and the wages of the workers. The second part is the firm's desired profit, which is determined, among other factors, by the profit margins in that particular business. If you add these two parts together, you get the price the firm wishes to charge. The quantity Q0 and associated price P0 give you one point on the firm's supply curve, as shown in Figure 10.

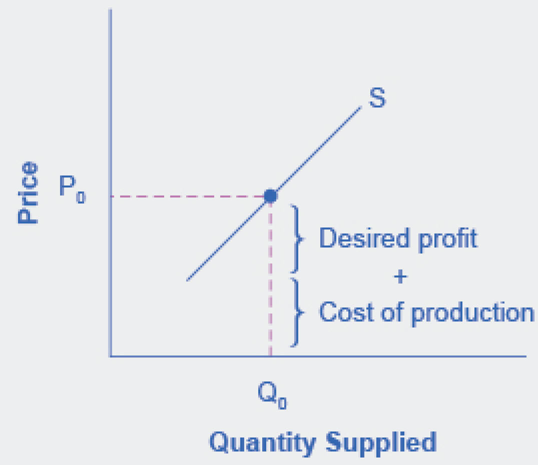


Figure 10: The cost of production and the desired profit equal the price a firm will set for a product.

Step 3. Now, suppose that the cost of production goes up. Perhaps cheese has become more expensive by R11.50 per pizza. If that is true, the firm will want to raise its price by the amount of the increase in cost (R11.50). Draw this point on the supply curve directly above the initial point on the curve, but R11.50 higher, as shown in Figure 11.

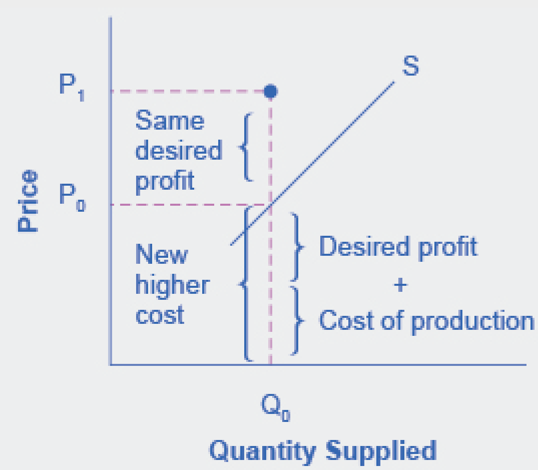


Figure 11: Because the cost of production and the

desired profit equal the price a firm will set for a product, if the cost of production increases, the price for the product will also need to increase.

Step 4. Shift the supply curve through this point. You will see that an increase in cost causes a leftward shift of the supply curve so that at any price, the quantities supplied will be smaller, as shown in Figure 12.

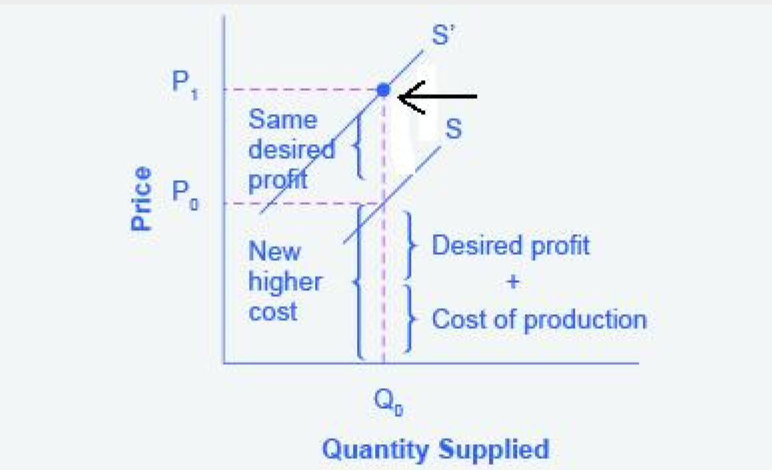


Figure 12: When the cost of production increases, the supply curve shifts left so that firms would have to charge a higher price to cover their higher costs to produce the same volume as before cost increased.

Summing Up: Factors That Change Supply

Changes in the cost of inputs, natural disasters, new technologies, and the impact of government decisions all affect the cost of production. In turn, these factors affect how much firms are willing to supply at any given price.

Figure 13 summarizes factors that change the supply of goods and services. Notice that a change in the price of the product itself is not among the factors that shift the supply curve. Although a change in price of a good or service typically causes a change in quantity supplied or a movement along the supply curve for that specific good or service, it does not cause the supply curve itself to shift.

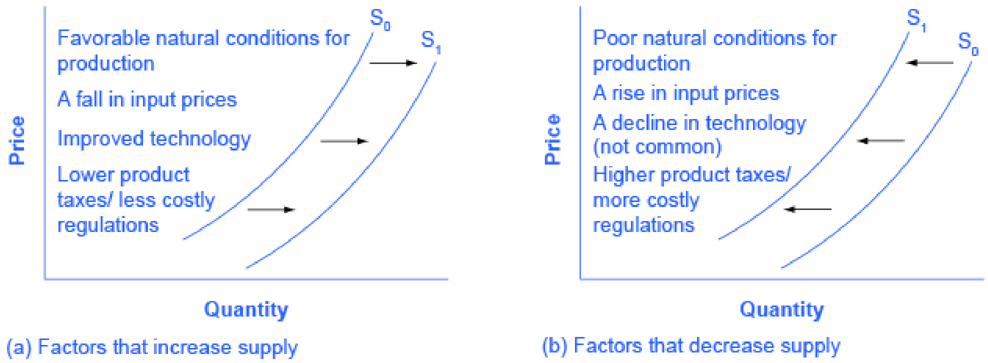


Figure 13: (a) A list of factors that can cause an increase in supply from S0 to S1. (b) The same factors, if their direction is reversed, can cause a decrease in supply from S0 to S1.

Because demand and supply curves appear on a two-dimensional diagram with only price and quantity on the axes, an unwary visitor to the land of economics might be fooled into believing that economics is about only four topics: demand, supply, price, and quantity. However, demand and supply are really “umbrella” concepts: demand covers all the factors that affect demand, and supply covers all the factors that affect supply. Factors other than price that affect demand and supply are included by using shifts in the demand or the supply curve. In this way, the two-dimensional demand and supply model becomes a powerful tool for analyzing a wide range of economic circumstances.

Key Concepts and Summary

Economists often use the *ceteris paribus* or “other things being equal” assumption: while examining the economic impact of one event, all other factors remain unchanged for the purpose of the analysis. Factors that can shift the demand curve for goods and services, causing a different quantity to be demanded at any given price, include changes in tastes, population, income, prices of substitute or complement goods, and expectations about future conditions and prices. Factors that can shift the supply curve for goods and services, causing a different quantity to be supplied at any given price, include input prices, natural conditions, changes in technology, and government taxes, regulations, or subsidies.

Self-Check Questions

Exercise:

Problem: Why do economists use the *ceteris paribus* assumption?

Solution:

To make it easier to analyze complex problems. *Ceteris paribus* allows you to look at the effect of one factor at a time on what it is you are trying to analyze. When you have analyzed all the factors individually, you add the results together to get the final answer.

Exercise:

Problem:

In an analysis of the market for paint, an economist discovers the facts listed below. State whether each of these changes will affect supply or demand, and in what direction.

a. There have recently been some important cost-saving inventions in the technology for making paint.
b. Paint is lasting longer, so that property owners need not repaint as often.
c. Because of severe hailstorms, many people need to repaint now.

d. The hailstorms damaged several factories that make paint, forcing them to close down for several months.

Solution:

a. An improvement in technology that reduces the cost of production will cause an increase in supply. Alternatively, you can think of this as a reduction in price necessary for firms to supply any quantity. Either way, this can be shown as a rightward (or downward) shift in the supply curve.
b. An improvement in product quality is treated as an increase in tastes or preferences, meaning consumers demand more paint at any price level, so demand increases or shifts to the right. If this seems a strange result, note that demand in the future for the longer-lasting paint will fall. Consumers are essentially shifting demand from the future to the present.
c. An increase in need causes an increase in demand or a rightward shift in the demand curve.
d. Factory damage means that firms are unable to supply as much in the present. Technically, this is an increase in the cost of production. Either way you look at it, the supply curve shifts to the left.

Exercise:

Problem:

Many changes are affecting the market for fuel oil. Predict how each of the following events will affect the equilibrium price and quantity in the market for oil. In each case, state how the event will affect the supply and demand diagram. Create a sketch of the diagram if necessary.

a. Cars are becoming more fuel efficient, and therefore travel more km per litre.
b. The winter is exceptionally cold.

c. A major discovery of new oil is made off the coast of South Africa.
d. The economies of some major oil-using nations, like Japan, slow down.
e. A war in the Middle East disrupts oil-pumping schedules.
f. The prices of electric heaters fall dramatically.
g. The price of solar energy falls dramatically.
h. Chemical companies invent a new, popular kind of plastic made from oil.

Solution:

a. More fuel-efficient cars means there is less need for petrol. This causes a leftward shift in the demand for petrol and thus oil. Since the demand curve is shifting down the supply curve, the equilibrium price and quantity both fall.
b. Cold weather increases the need for heating oil. This causes a rightward shift in the demand for heating oil and thus oil. Since the demand curve is shifting up the supply curve, the equilibrium price and quantity both rise.
c. A discovery of new oil will make oil more abundant. This can be shown as a rightward shift in the supply curve, which will cause a decrease in the equilibrium price along with an increase in the equilibrium quantity. (The supply curve shifts down the demand curve so price and quantity follow the law of demand. If price goes down, then the quantity goes up.)
d. When an economy slows down, it produces less output and demands less input, including energy, which is used in the production of virtually everything. A decrease in demand for energy will be reflected as a decrease in the demand for oil, or a leftward shift in demand for oil. Since the demand curve is shifting down the supply curve, both the equilibrium price and quantity of oil will fall.
e. Disruption of oil pumping will reduce the supply of oil. This leftward shift in the supply curve will show a movement up the

demand curve, resulting in an increase in the equilibrium price of oil and a decrease in the equilibrium quantity.

f. A decrease in the price of electric heaters will lead to an increase in quantity demanded of electric heaters and decrease in demand for oil heaters and decrease in demand for oil. This leftward shift in the demand for oil causes a movement down the supply curve, resulting in a decrease in the equilibrium price and quantity of oil.
g. Solar energy is a cleaner substitute for oil-based energy. So if solar energy becomes cheaper, the demand for oil will decrease as consumers switch from oil to solar. The decrease in demand for oil will be shown as a leftward shift in the demand curve. As the demand curve shifts down the supply curve, both equilibrium price and quantity for oil will fall.
h. A new, popular kind of plastic will increase the demand for oil. The increase in demand will be shown as a rightward shift in demand, raising the equilibrium price and quantity of oil.

Review Questions

Exercise:

Problem:

When analyzing a market, how do economists deal with the problem that many factors that affect the market are changing at the same time?

Exercise:

Problem:

Name some factors that can cause a shift in the demand curve in markets for goods and services.

Exercise:

Problem:

Name some factors that can cause a shift in the supply curve in markets for goods and services.

Critical Thinking Questions

Exercise:

Problem:

Consider the demand for burgers. If the price of a substitute good (for example, hot dogs) increases and the price of a complement good (for example, burger buns) increases, can you tell for sure what will happen to the demand for burgers? Why or why not? Illustrate your answer with a sketch graph.

Exercise:

Problem:

We know that a change in the price of a product causes a movement along the demand curve. Suppose consumers believe that prices will be rising in the future. How will that affect demand for the product in the present? Can you show this graphically?

Exercise:

Problem:

The South African government proposes to introduce a sugar tax. How would this affect the supply of fizzy cold drinks and their equilibrium price and quantity? Can you show this graphically? *Hint*: assume that the sugar tax is collected from the sellers

Problems

Exercise 10

The table below shows the supply and demand information for bicycles measured in thousands per year.

Price (R/bicycle)	Qd (bicycles0	Qs (bicycles)
1800	50	36
2250	40	40
2700	32	48
3150	28	56
3600	24	70

Table 3. Supply and demand information for bicycles.

- What is the quantity demanded and the quantity supplied at a price of R210?
- At what price is the quantity supplied equal to 48,000?
- Graph the demand and supply curve for bicycles. How can you determine the equilibrium price and quantity from the graph? How can you determine the equilibrium price and quantity from the table? What are the equilibrium price and equilibrium quantity?
- If the price was R120, what would the quantities demanded and supplied be? Would a shortage or surplus exist? If so, how large would the shortage or surplus be?

Exercise:

Problem:

The computer market in recent years has seen many more computers sell at much lower prices. What shift in demand or supply is most likely to explain this outcome? Sketch a demand and supply diagram and explain your reasoning for each.

a. A rise in demand
b. A fall in demand
c. A rise in supply
d. A fall in supply

References

Burgin, R. 2015. Case Study: South Africa’s growing poultry consumption. World Poultry,17 Dec. Available: http://www.worldpoultry.net/Meat/Articles/2015/12/Case-Study-South-Africas-growing-poultry-consumption-2735530W/ (Accessed: 23 Apr 2016)

Vollgraaf, R. and Mbatha, A. 2016. South African growth weakens amid worst drought in 100 years. Bloomberg news, 1 Mar. Available: http://www.bloomberg.com/news/articles/2016-03-01/south-african-gdp-growth-slows-to-0-6-in-fourth-quarter (Accessed: 23 Apr 2016)

World Bank. 2015. South Africa’s changing demographic could lift growth by 5.4% by 2030. Press release, 17 Aug. Available: http://www.worldbank.org/en/news/press-release/2015/08/17/south-africa-demographic-lift-growth (Accessed: 23 Apr 2016)

Glossary

ceteris paribus
: other things being equal

complements
: goods that are often used together so that consumption of one good tends to enhance consumption of the other

factors of production
: the combination of labor, materials, and machinery that is used to produce goods and services; also called inputs

inferior good
: a good in which the quantity demanded falls as income rises, and in which quantity demanded rises and income falls

inputs
: the combination of labor, materials, and machinery that is used to produce goods and services; also called factors of production

normal good

a good in which the quantity demanded rises as income rises, and in which quantity demanded falls as income falls

shift in demand
: when a change in some economic factor (other than price) causes a different quantity to be demanded at every price

shift in supply
: when a change in some economic factor (other than price) causes a different quantity to be supplied at every price

substitute
: a good that can replace another to some extent, so that greater consumption of one good can mean less of the other

Changes in Equilibrium Price and Quantity: The Four-Step Process

By the end of this section, you will be able to:

- Identify equilibrium price and quantity through the four-step process
- Graph equilibrium price and quantity
- Contrast shifts of demand or supply and movements along a demand or supply curve
- Graph demand and supply curves, including equilibrium price and quantity, based on real-world examples

Figure 1

Let’s begin this discussion with a single economic event. It might be an event that affects demand, like a change in income, population, tastes, prices of substitutes or complements, or expectations about future prices. It might be an event that affects supply, like a change in natural conditions, input prices, or technology, or government policies that affect production.

How does this economic event affect equilibrium price and quantity? We will analyze this question using a four-step process.

Step 1. Draw a demand and supply model before the economic change took place. To establish the model requires four standard pieces of information: The law of demand, which tells us the slope of the demand curve; the law of supply, which gives us the slope of the supply curve; the shift variables for demand; and the shift variables for supply. From this model, find the initial equilibrium values for price and quantity.

Step 2. Decide whether the economic change being analyzed affects demand or supply. In other words, does the event refer to something in the list of demand factors or supply factors?

Step 3. Decide whether the effect on demand or supply causes the curve to shift to the right or to the left, and sketch the new demand or supply curve on the diagram. In other words, does the event increase or decrease the amount consumers want to buy or producers want to sell?

Step 4. Identify the new equilibrium and then compare the original equilibrium price and quantity to the new equilibrium price and quantity.

Let’s consider one example that involves a shift in supply and one that involves a shift in demand. Then we will consider an example where both supply and demand shift.

Worst drought in 100 years for SA farmers

South Africa's Department of Agriculture, Forestry and Fisheries (2015) reported that the average producer price of bananas increased by 85% between the fourth quarter of 2013 and the fourth quarter of 2014. This large increase in price was attributed to a decrease in local production which, in turn, might be due at least partly to the worsening drought conditions South Africa experienced from about 2013. What is the economic explanation of how these climate conditions affected the quantity and price of bananas? Figure 2 illustrates the four-step approach, which is

explained below, to work through this problem. Table 1 provides the information to work the problem as well.

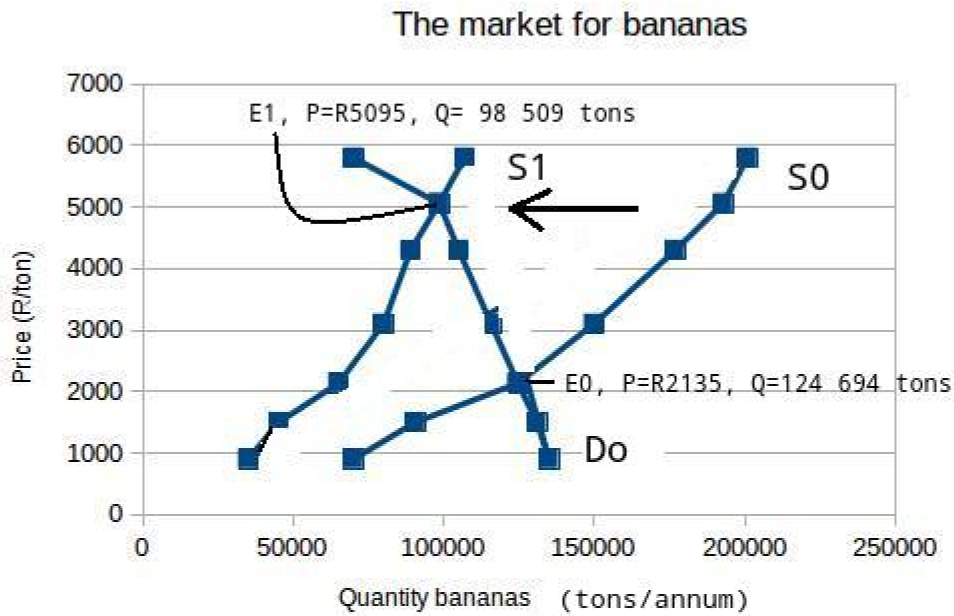


Figure 2. Climate change has resulted in a change in the price and quantity of bananas traded

Price (R/ton)	Supply in 2013 (tons)	Supply in 2014 (tons)	Demand (tons)
900	70 000	35 000	135 000
1 500	90 600	45 200	131 000
2 135	124 694	65 000	124 694
3 100	150 000	80 000	115 800
4 300	177 000	89 000	105 000
5 059	193 000	98 509	98 509
5 800	201 000	107 000	70 100

Table 1. Market for bananas: price and quantity

Step 1. Draw a demand and supply model to illustrate the market for bananas in the year before the drought conditions worsened in 2013. The demand curve D_0 and the supply curve S_0 show that the original equilibrium price is R2135 per ton and the original equilibrium quantity is 124 694 tons.

Step 2. Did the economic event affect supply or demand? Bad weather or climate is an example of a natural condition that affects supply.

Step 3. Was the effect on supply an increase or a decrease? Bad weather or climate is a change in natural conditions that decreases the quantity supplied at any given price. The supply curve shifts to the left, moving from the original supply curve S0 to the new supply curve S1, which is shown in both the table and the figure.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium. At the new equilibrium E1, the equilibrium price has increased to R5059 from R2135 and the equilibrium quantity decreased from 124 694 tons to 98 509 tons. Notice that the equilibrium quantity demanded decreased, even though the demand curve did not move.

In short, drought conditions decreased the supply of bananas. The result was a lower equilibrium quantity of bananas bought and sold in the market at a higher price. How did price increase? Recall that temporary shortages will force price higher. With the reduced supply following the drought, price could not remain at R2135 per ton as this would result in a shortage. And so this shortage forces price up gradually reducing the shortage through the laws of demand and supply until it has been eliminated. Once there is no longer a shortage, price comes to rest at a level where quantity demanded exactly matches the new quantity supplied (equilibrium).

Newspapers and the Internet

According to the Pew Research Center (Mitchell: 2015), more and more people, especially younger people, are getting their news from online and digital sources. The majority of U.S. adults now own smartphones or tablets, and most of those Americans say they use them in part to get the

news. The worldwide trend of declining newspaper sales due to the growth of online media is also a reality in South Africa which is experiencing falling sales of printed newspapers (Mokgata: 2013). Let's use the demand and supply model to figure out what's going on.

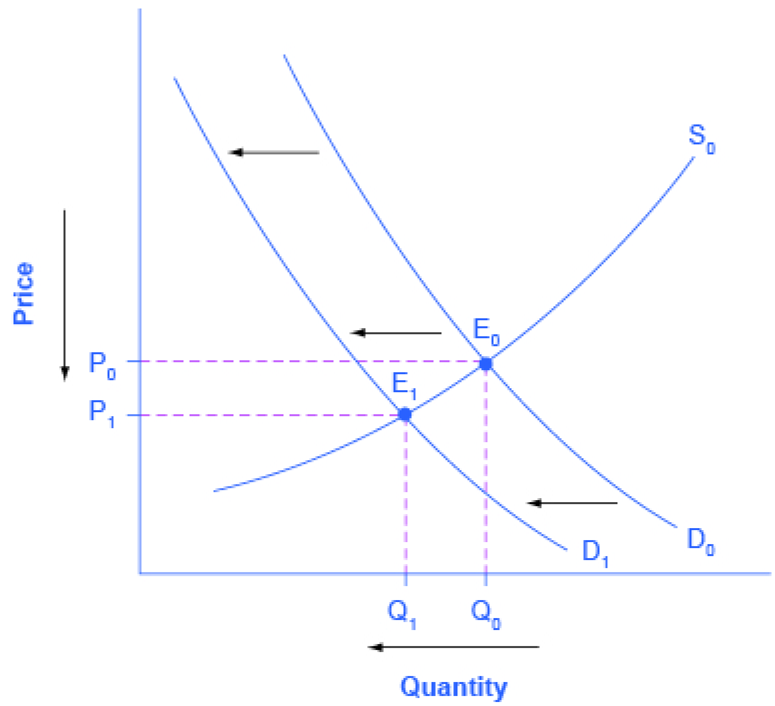


Figure 3: A change in tastes from print news sources to digital sources results in a leftward shift in demand for print newspapers. The result is a decrease in both equilibrium price and quantity.

Step 1. Develop a demand and supply model to think about what the market looked like before the event. The demand curve D_0 and the supply curve S_0 show the original relationships. In this case, the analysis is performed without specific numbers on the price and quantity axis.

Step 2. Did the change described affect supply or demand? A change in tastes, from traditional news sources (print, radio, and television) to digital

sources, caused a change in demand for the former.

Step 3. Was the effect on demand positive or negative? A shift to digital news sources will tend to mean a lower quantity demanded of traditional news sources at every given price, causing the demand curve for print and other traditional news sources to shift to the left, from D_0 to D_1.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_1) occurs at a lower quantity and a lower price than the original equilibrium (E_0). How is it that publishers find they must now reduce prices to sell news print? Well, because printed newspapers are now less popular or fashionable (so less demand), there would be a surplus of them at the old higher prices given that there is less demand for these products. This temporary surplus will cause prices to fall and so the laws of demand and supply will work to gradually eliminate the surplus as price falls. Price will come to rest once the surplus has been eliminated at the output level where quantity demanded and quantity supplied exactly match (equilibrium).

The Interconnections and Speed of Adjustment in Real Markets

In the real world, many factors that affect demand and supply can change all at once. For example, the demand for cars might increase because of rising incomes and population, and it might decrease because of rising fuel prices (a complementary good). Likewise, the supply of cars might increase because of innovative new technologies that reduce the cost of car production, and it might decrease as a result of new government regulations requiring the installation of costly pollution-control technology.

Moreover, rising incomes and population or changes in fuel prices will affect many markets, not just cars. How can an economist sort out all these interconnected events? The answer lies in the *ceteris paribus* assumption. Look at how each economic event affects each market, one event at a time, holding all else constant. Then combine the analyses to see the net effect.

A Combined Example

The South African Postal Service is facing difficult challenges due to falling mail volumes and past labor strikes (Speckman: 2013). Compensation for postal workers tends to increase most years due to cost-of-living increases. At the same time, more and more people are using email, text, and other digital message forms such as Facebook and Twitter to communicate with friends and others. What does this suggest about the continued viability of the Postal Service? Figure 4 and the text below illustrates using the four-step analysis to answer this question.

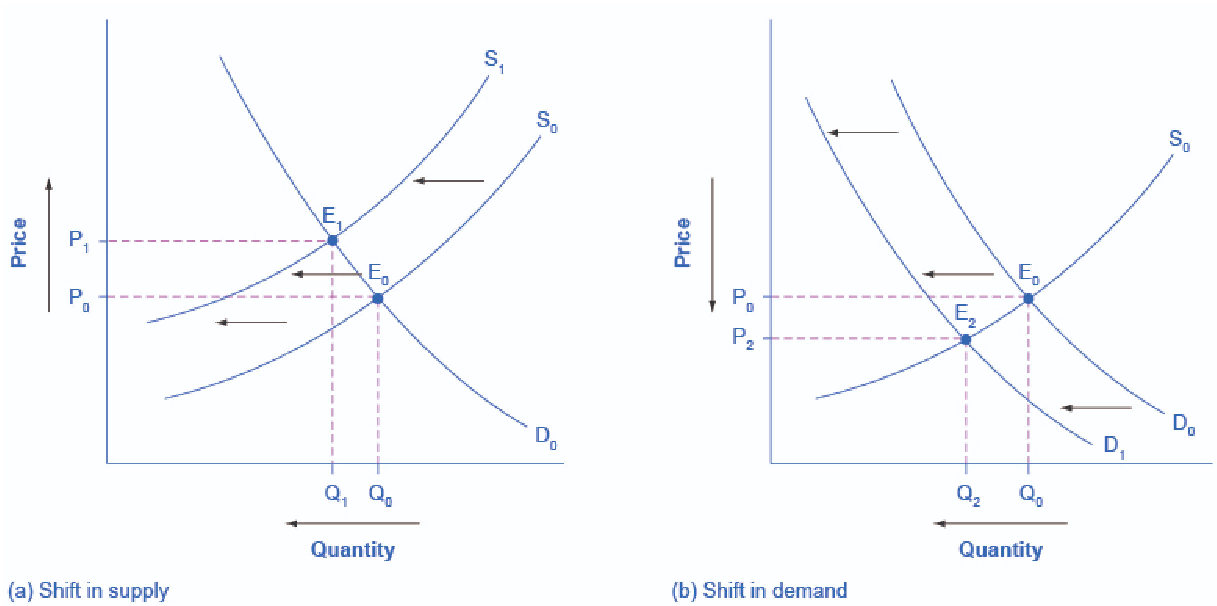


Figure 4: (a) Higher labor compensation causes a leftward shift in the supply curve, a decrease in the equilibrium quantity, and an increase in the equilibrium price. (b) A change in tastes away from Postal Services causes a leftward shift in the demand curve, a decrease in the equilibrium quantity, and a decrease in the equilibrium price.

Since this problem involves two disturbances, we need two four-step analyses, the first to analyze the effects of higher compensation for postal workers, the second to analyze the effects of many people switching from "snailmail" to email and other digital messages.

Figure 4 (a) shows the shift in supply discussed in the following steps.

Step 1. Draw a demand and supply model to illustrate what the market for the South African Postal Service looked like before this scenario starts. The demand curve D_0 and the supply curve S_0 show the original relationships.

Step 2. Did the change described affect supply or demand? Labor compensation is a cost of production. A change in production costs caused a change in supply for the Postal Service.

Step 3. Was the effect on supply positive or negative? Higher labor compensation leads to a lower quantity supplied of postal services at every given price, causing the supply curve for postal services to shift to the left, from S_0 to S_1.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_1) occurs at a lower quantity and a higher price than the original equilibrium (E_0).

Figure 3 (b) shows the shift in demand discussed in the following steps.

Step 1. Draw a demand and supply model to illustrate what the market for South African Postal Services looked like before this scenario starts. The demand curve D_0 and the supply curve S_0 show the original relationships. Note that this diagram is independent from the diagram in panel (a).

Step 2. Did the change described affect supply or demand? A change in tastes away from snailmail toward digital messages will cause a change in demand for the Postal Service.

Step 3. Was the effect on supply positive or negative? Higher labor compensation leads to a lower quantity supplied of postal services at every

given price, causing the supply curve for postal services to shift to the left, from D_0 to D_1.

Step 4. Compare the new equilibrium price and quantity to the original equilibrium price. The new equilibrium (E_2) occurs at a lower quantity and a lower price than the original equilibrium (E_0).

The final step in a scenario where both supply and demand shift is to combine the two individual analyses to determine what happens to the equilibrium quantity and price. Graphically, we place the previous two diagrams one on top of the other, as in Figure 5.

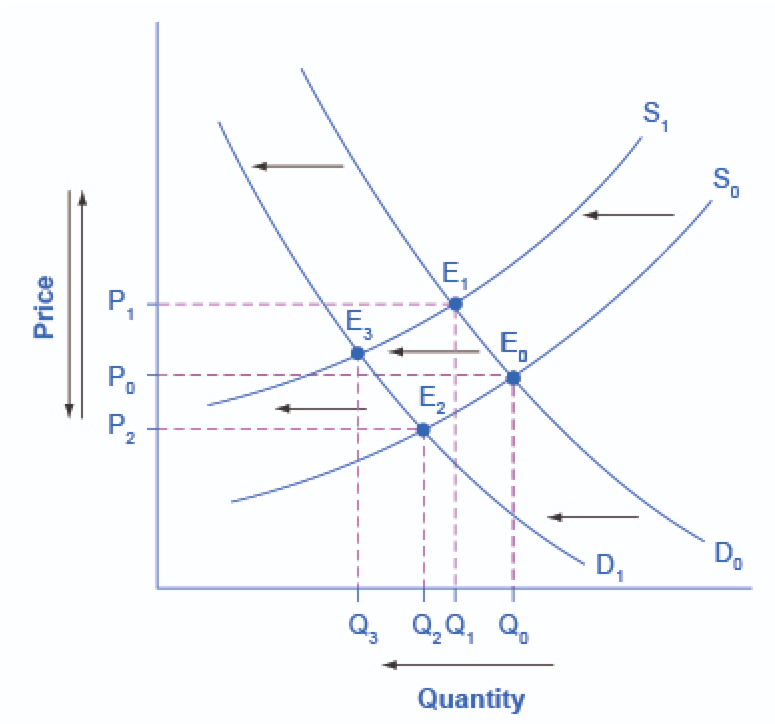


Figure 5: Supply and demand shifts cause changes in equilibrium price and quantity.

The following are the results:

Effect on Quantity: The effect of higher labor compensation on Postal Services is to decrease the equilibrium quantity because it raises the cost of production. The effect of a change in tastes away from posted mail is to decrease the equilibrium quantity. Since both shifts are to the left, the overall impact is a decrease in the equilibrium quantity of Postal Services (Q_3). This is easy to see graphically, since Q_3 is to the left of Q_0.

Effect on Price: The overall effect on price is more complicated. The effect of higher labor compensation on Postal Services is to increase the equilibrium price, because it raises the cost of production. The effect of a change in tastes away from posted mail is to decrease the equilibrium price. Since the two effects are in opposite directions, unless we know the sizes of the two effects, the overall effect is unclear. This is not unusual. When both curves shift, typically we can work out the overall effect on price, or on quantity, but not on both. In this case, we worked out the overall effect on the equilibrium quantity, but not on the equilibrium price. In other cases, it might be the opposite.

The next Clear It Up feature focuses on the difference between shifts of supply or demand and movements along a curve.

Note:

What is the difference between shifts of demand or supply versus movements along a demand or supply curve?

One common mistake in applying the demand and supply model is to confuse the shift of a demand or a supply curve with movement along a demand or supply curve (a change in quantity demanded or quantity supplied). As an example, consider a problem that asks whether a drought will increase or decrease the equilibrium quantity and equilibrium price of wheat. Mbongeni, a student in an introductory economics class, might argue as follows:

“Well, it is clear that a drought reduces supply, so I will shift back the supply curve, as in the shift from the original supply curve S_0 to S_1 shown on the diagram (called Shift 1). So the equilibrium moves from E_0 to E_1, the equilibrium quantity is lower and the equilibrium price is higher. Then, a higher price makes farmers more likely to supply the good, so the supply

curve shifts right, as shown by the shift from S_1 to S_2, on the diagram (shown as Shift 2), so that the equilibrium now moves from E_1 to E_2. The higher price, however, also reduces demand and so causes demand to shift back, like the shift from the original demand curve, D_0 to D_1 on the diagram (labeled Shift 3), and the equilibrium moves from E_2 to E_3."

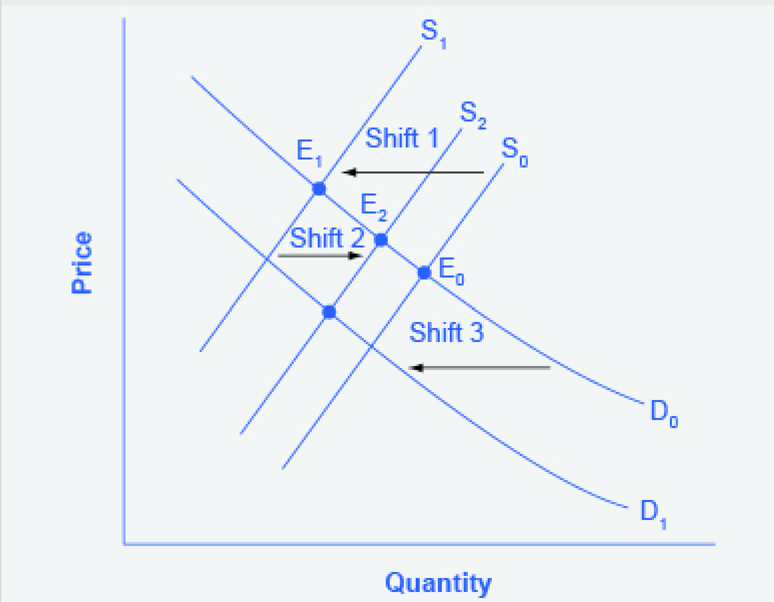


Figure 6: A shift in one curve never causes a shift in the other curve. Rather, a shift in one curve causes a movement along the second curve.

At about this point, Mbongeni suspects that this answer is headed down the wrong path. Think about what might be wrong with his argument, and then read the answer that follows.
Answer: Mbongeni's first step is correct: that is, a drought shifts back the supply curve of wheat and leads to a prediction of a lower equilibrium quantity and a higher equilibrium price. This corresponds to a movement along the original demand curve (D_0), from E_0 to E_1. The rest of Mbongeni's argument is wrong, because it mixes up shifts in supply with quantity supplied, and shifts in demand with quantity demanded. A higher or lower price never shifts the supply curve, as suggested by the shift in

supply from S_1 to S_2. Instead, a price change leads to a movement along a given supply curve. Similarly, a higher or lower price never shifts a demand curve, as suggested in the shift from D_0 to D_1. Instead, a price change leads to a movement along a given demand curve. Remember, a change in the price of a good never causes the demand or supply curve for that good to shift.

Think carefully about the timeline of events: What happens first, what happens next? What is cause, what is effect? If you keep the order right, you are more likely to get the analysis correct.

In the four-step analysis of how economic events affect equilibrium price and quantity, the movement from the old to the new equilibrium seems immediate. As a practical matter, however, prices and quantities often do not zoom straight to equilibrium. More realistically, when an economic event causes demand or supply to shift, prices and quantities set off in the general direction of equilibrium. Indeed, even as they are moving toward one new equilibrium, prices are often then pushed by another change in demand or supply toward another equilibrium.

Key Concepts and Summary

When using the supply and demand model to think about how an event will affect the equilibrium price and quantity, proceed through four steps: (1) sketch a supply and demand diagram to think about what the market looked like before the event; (2) decide whether the event will affect supply or demand; (3) decide whether the effect on supply or demand is negative or positive, and draw the appropriate shifted supply or demand curve; (4) compare the new equilibrium price and quantity to the original ones.

Self-Check Questions

Exercise:

Problem:

Let’s think about the market for air travel. From August 2014 to January 2015, the price of jet fuel decreased roughly 47%. Using the four-step analysis, how do you think this fuel price decrease affected the equilibrium price and quantity of air travel?

Solution:

Step 1. Draw the graph with the initial supply and demand curves. Label the initial equilibrium price and quantity.

Step 2. Did the economic event affect supply or demand? Jet fuel is a cost of producing air travel, so an increase in jet fuel price affects supply.

Step 3. An increase in the price of jet fuel caused a decrease in the cost of air travel. We show this as a downward or rightward shift in supply.

Step 4. A rightward shift in supply causes a movement down the demand curve, lowering the equilibrium price of air travel and increasing the equilibrium quantity.

Exercise:

Problem:

A tariff is a tax on imported goods. Suppose the South African government cuts the tariff on imported flat screen televisions. Using the four-step analysis, how do you think the tariff reduction will affect the equilibrium price and quantity of flat screen TVs?

Solution:

Step 1. Draw the graph with the initial supply and demand curves. Label the initial equilibrium price and quantity.

Step 2. Did the economic event affect supply or demand? A tariff is treated like a cost of production, so this affects supply.

Step 3. A tariff reduction is equivalent to a decrease in the cost of production, which we can show as a rightward (or downward) shift in supply.

Step 4. A rightward shift in supply causes a movement down the demand curve, lowering the equilibrium price and raising the equilibrium quantity.

Review Questions

Exercise:

Problem:

How does one analyze a market where both demand and supply shift?

Exercise:

Problem:

What causes a movement along the demand curve? What causes a movement along the supply curve?

Critical Thinking Questions

Exercise:

Problem:

Use the four-step process to analyze the impact of Uber taxis (a cheap and safe substitute for normal taxis) on the market for normal taxi rides.

Exercise:

Problem:

Use the four-step process to analyze the impact of government regulation and licensing requirements imposed on Uber taxis on the market for normal taxi rides.

Exercise:

Problem:

Show graphically the effect on the market for normal taxi rides of the introduction of Uber taxis as well as that many traditional taxi drivers have closed their businesses due to the competition from Uber taxis (both events happening at the same time).

Problems

Exercise:

Problem:

Demand and supply in the market for cheddar cheese is illustrated in Table 2. Graph the data and find the equilibrium. Next, create a table showing the change in quantity demanded or quantity supplied, and a graph of the new equilibrium, in each of the following situations:

a. The price of milk, a key input for cheese production, rises, so that the supply decreases by 80 Kg at every price.
b. A new study says that eating cheese is good for your health, so that demand increases by 20% at every price.

Price per Kg	**Qd**	**Qs**
R45.00	750	540
R48.00	700	600
R51.00	**650**	**650**
R54.00	620	700
R57.00	600	720
R60.00	590	730

Table 2. Demand for, and supply of, cheddar

cheese.

References

Mitchell, A. 2015. State of the news media 2015. Pew Research Center 29 Apr. Available: http://www.journalism.org/2015/04/29/state-of-the-news-media-2015/ (Accessed: 23 Apr 2016)

Mokgata, Z. 2013. SA's newspapers "under pressure" as circulation drops. Business Day Live, 14 Aug. Available: http://www.bdlive.co.za/national/media/2013/08/14/sas-newspapers-under-pressure-as-circulation-drops (Accessed: 23 Apr 2016)

South Africa. Department of Agriculture, Forestry and Fisheries. 2015. Banana. Market analysis report no. 1 of 2015. Pretoria: Department of Agriculture, Forestry and Fisheries. Available: http://www.nda.agric.za/doaDev/sideMenu/Marketing/Quartely%20Bulletins%20Uploaded/Banana%20Market%20analysis%20Report%20No%20%201%20of%202015.pdf (Accessed: 31 Oct 2016)

Speckman, A. 2013. SA Post Office faces big challenges. IOL news, 19 Sep. Available: http://www.iol.co.za/business/companies/sa-post-office-faces-big-challenges-1579654 (Accessed: 23 Apr 2016)

Price Ceilings, Price Floors And A Look At The Labor Market

By the end of this section, you will be able to:

- Explain price controls, price ceilings, and price floors
- Analyze demand and supply as a social adjustment mechanism

Controversy sometimes surrounds the prices and quantities established by demand and supply, especially for products that are considered necessities. In some cases, discontent over prices turns into public pressure on politicians, who may then pass legislation to prevent a certain price from climbing “too high” or falling “too low.”

The demand and supply model shows how people and firms will react to the incentives provided by these laws to control prices, in ways that will often lead to undesirable consequences. Alternative policy tools can often achieve the desired goals of price control laws, while avoiding at least some of their costs and tradeoffs.

Price Ceilings

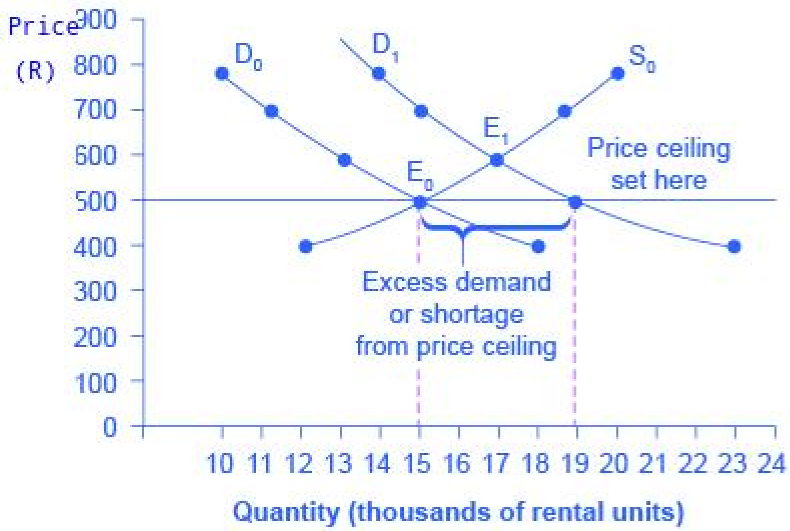
Laws that government enacts to regulate prices are called **Price controls**. Price controls come in two flavors. A **price ceiling** keeps a price from rising above a certain level (the “ceiling”), while a **price floor** keeps a price from falling below a certain level (the “floor”). This section uses the demand and supply model to analyze price ceilings. To be effective a price ceiling must be set above the current market equilibrium price. The next section discusses price floors.

In many markets for goods and services, 'demanders' outnumber suppliers. Consumers, who are also potential voters, sometimes unite behind a political proposal to hold down a certain price. South Africa's Rental Housing Act 50 of 1999 gives the Rental Housing Tribunals in each province the power to determine rentals that are just and equitable to parties by considering several factors outlined in the Act (Mohamed: 2012). A ruling by a Rental Housing Tribunal in KwaZulu-Natal, for example, would

amount to a price ceiling. What effect would this have on the rental housing market?

Rent control becomes a politically hot topic when rents begin to rise rapidly. Everyone needs an affordable place to live. Perhaps a change in tastes makes a certain suburb or town a more popular place to live. Perhaps locally-based businesses expand, bringing higher incomes and more people into the area. Changes of this sort can cause a change in the demand for rental housing, as Figure 1 illustrates. The original equilibrium (E_0) lies at the intersection of supply curve S_0 and demand curve D_0, corresponding to an equilibrium price of R5000 and an equilibrium quantity of 15,000 units of rental housing. The effect of greater income or a change in tastes is to shift the demand curve for rental housing to the right, as shown by the data in Table 1 and the shift from D_0 to D_1 on the graph. In this market, at the new equilibrium E_1, the price of a rental unit would rise to R6000 and the equilibrium quantity would increase to 17,000 units.

A Price Ceiling Example—Rent Control



The original intersection of demand and supply occurs at E_0. If demand shifts from D_0 to D_1, the new equilibrium would be at E_1—unless a price ceiling prevents the price from rising. If the price is not permitted to rise, the quantity supplied remains at 15,000. However, after the

change in demand, the quantity demanded rises to 19,000, resulting in a shortage.

Price (R/month)	Original Quantity Supplied (units)	Original Quantity Demanded (units)	New Quantity Demanded (units)
R4000.00	12,000	18,000	23,000
R5000.00	15,000	15,000	19,000
R6000.00	17,000	13,000	17,000
R7000.00	19,000	11,000	15,000
R8000.00	20,000	10,000	14,000

Rent Control

Suppose that the Rental Housing Tribunal rules that monthly price of rented accommodation may not exceed the original equilibrium of R5000 for a typical 3-bedroom apartment. In Figure 1, the horizontal line at the price of R5000 shows the legally fixed maximum price set by the rent control law. However, the underlying forces that shifted the demand curve to the right are still there. At that price (R5000), the quantity supplied remains at the same 15,000 rental units, but the quantity demanded is 19,000 rental units. In other words, the quantity demanded exceeds the quantity supplied, so there is a permanent shortage of rental housing (since price may not adjust upward due to the shortage as would normally happen). One of the

unintended consequences of price ceilings is that while the price ceiling was intended to help renters, there are actually fewer apartments rented out under the price ceiling (15,000 rental units) than would be the case at the market rent of R6000 (17,000 rental units).

Price ceilings do not simply benefit renters at the expense of landlords. Rather, some renters (or potential renters) lose their housing as landlords convert apartments to alternative more profitable uses (e.g.offices). Even when the housing remains in the rental market, landlords tend to spend less on maintenance and on essentials like heating, cooling, hot water, and lighting since they cannot increase the rental to pay for these improvements. The first rule of economics is you do not get something for nothing—everything has an opportunity cost. So if renters get “cheaper” housing than the market requires, they tend to also end up with lower quality housing.

Price ceilings have been proposed for other products. For example, price ceilings to limit what producers can charge have been proposed in recent years for prescription medicines, doctor and hospital fees, the charges made by some automatic teller bank machines, and car insurance rates. Price ceilings are enacted in an attempt to keep prices low for those who demand the product (to protect consumers). But when the market price is not allowed to rise to the equilibrium level, quantity demanded exceeds quantity supplied, and thus a permanent shortage occurs. Those who manage to purchase the product at the lower price given by the price ceiling will benefit, but sellers of the product will suffer, along with those who are not able to purchase the product at all. Quality is also likely to deteriorate.

Price Floors

A price floor is the lowest legal price that can be paid in markets for goods and services or even in some cases resources such as labor. To be effective it must be set above the current market equilibrium price. Perhaps the best-known example of a price floor is the minimum wage, which is based on the normative view that someone working full time ought to be able to afford a basic standard of living. Minimum wage legislation applies to most sectors in South Africa. Different minimum wages are prescribed annually by government for the different sectors. So, for example, the minimum

wage determined for contract cleaners for the period 1-12-2015 to 30-11-2016 in KwaZulu-Natal was R16.41 per hour (Minimum wages in South Africa: 2016). This gives a monthly income for a single person slightly higher than the national poverty line which is the minimum monthly income a person requires to meet his or her basic food and essential requirements. South Africa has three poverty datum lines of which the second lowest (lower bound poverty line) poverty line in 2014 was set at R544 per person per month (Nicolson: 2015). As the cost of living rises over time, the government periodically raises the minimum wage.

Price floors are sometimes called “price supports,” because they support a price by preventing it from falling below a certain level. Around the world, many countries have passed laws to create agricultural price supports. Farm prices and thus farm incomes fluctuate, sometimes widely which causes uncertainty and hardship for farmers. So even if, on average, farm incomes are adequate, some years they can be quite low due to weather disruptions (such as drought, flood, hail, frost etc). The purpose of price supports is to prevent these large swings in farmers' incomes.

The most common way price supports work is that the government enters the market and buys up the product, adding to demand to keep prices higher than they otherwise would be. The South African government does not give price support protection (price floors) to local farmers. However, some other countries do. According to the Common Agricultural Policy reform passed in 2013, the European Union (EU) will spend about 60 billion euros per year, or 67 billion dollars per year, or roughly 38% of the EU budget, on price supports for Europe’s farmers from 2014 to 2020.

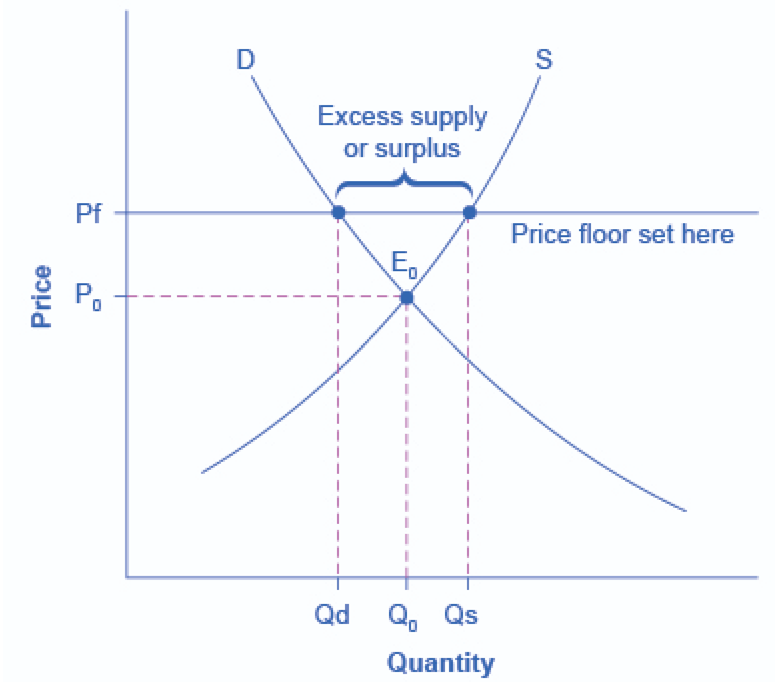
Figure 2 illustrates the effects of a government program that assures a price above the equilibrium by focusing on the market for wheat in Europe. In the absence of government intervention, the price would adjust so that the quantity supplied would equal the quantity demanded at the equilibrium point E_0, with price P_0 and quantity Q_0. However, policies to keep prices high for farmers keeps the price above what would have been the market equilibrium level—the price Pf shown by the dashed horizontal line in the diagram. The result is a quantity supplied in excess of the quantity

demanded (Qd). When quantity supplied exceeds quantity demanded, a surplus exists.

The high-income areas of the world, including the United States, Europe, and Japan, are estimated to spend roughly $1 billion per day in supporting their farmers. If the government is willing to purchase the excess supply (or to provide payments for others to purchase it), then farmers will benefit from the price floor, but taxpayers and consumers of food will pay the costs. Numerous proposals have been offered for reducing farm subsidies. In many countries, however, political support for subsidies for farmers remains strong. Either because this is viewed by the population as supporting the traditional rural way of life or because of the voting power of farmers and agricultural industries.

For more detail on the effects price ceilings and floors have on demand and supply, see the following Clear It Up feature.

European Wheat Prices: A Price Floor Example



The intersection of demand (D) and supply (S) would be at the equilibrium point E_0. However, a price floor set at Pf holds the price above E_0 and prevents it from falling.

The result of the price floor is that the quantity supplied Qs exceeds the quantity demanded Qd. There is excess supply, also called a surplus.

Note:
Do price ceilings and floors change demand or supply?
Neither price ceilings nor price floors cause demand or supply to change. They simply set a price that limits what can be legally charged in the market. Remember, changes in price do not cause demand or supply to change. Price ceilings and price floors can cause a different choice of quantity demanded along a demand curve, but they do not move the demand curve. Price controls can cause a different choice of quantity supplied along a supply curve, but they do not shift the supply curve.

Demand, Supply and Equilibrium in the Labor Market

Market supply of labor

The labor market is a link between potential sellers of labor services (supply of labor) and potential purchasers (demand for labor) of labor services. It is one of the resource markets. We know that full employment is a macroeconomic objective. The phenomenon of employment/unemployment can be understood in the context of the labor market i.e. unemployment can be explained using the supply and demand model.

The market supply and market demand together determine the equilibrium wage rate and employment levels. The market supply is simply the sum of all workers willing to offer their services at various wage rates. The market supply slopes positively. So an increase in wage rate will cause more people to offer their services. If we were to sketch the market supply of labor it would look like this in Figure 3:

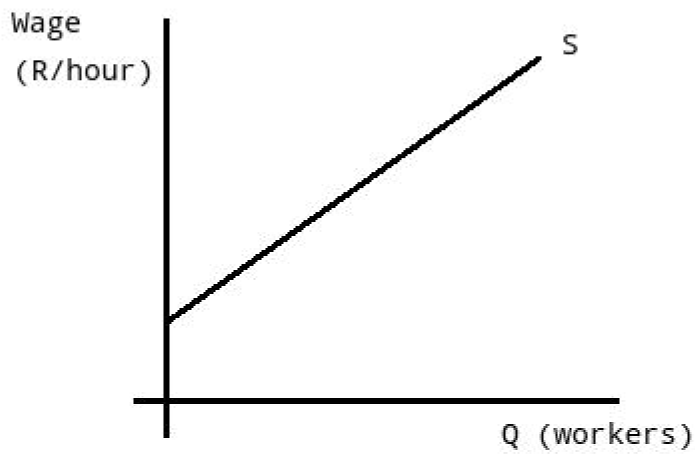


Figure 3. Market supply of labor.

Besides wages, there are non-wage determinants (ceteris paribus factors) of labor supply that will cause the entire labor market supply curve to shift. These include factors such as numbers of workers, training requirements, wages in other occupations, non-monetary aspects (e.g. safety, job security, work environment etc.) and trade union activity. Changes in any one or more of these factors will either cause labor market supply to decrease (shift left) or increase (shift right). A change in wage rate, on the other hand, will not cause supply to shift. It will result only in a change in the quantity of labor supplied (a movement along the supply curve).

Market demand for labor
Firms' demand for labor is a derived demand in the sense that it is derived from consumers' demand for goods and services produced by firms. At higher wage rates all firms will demand a smaller quantity of labor and more labor will be demanded at lower wage rates as shown in the sketch diagram:

Thus a change in wage rate will cause a change in the quantity of labor demanded by a firm (no shift of labor demand). However there are non-wage determinants (ceteris paribus factors) of labor demand that will cause

the market demand curve for labor to shift. These include a change in the number of firms requiring labor, a change in the demand for the firms' products, a change in labor productivity, a change in the price of substitutes for labor and a change in the cost of hiring labor. Any change in one or more of these ceteris paribus factors affecting a firms' demand for labor will either increase labor demand (shift right) or decrease it (shift left).

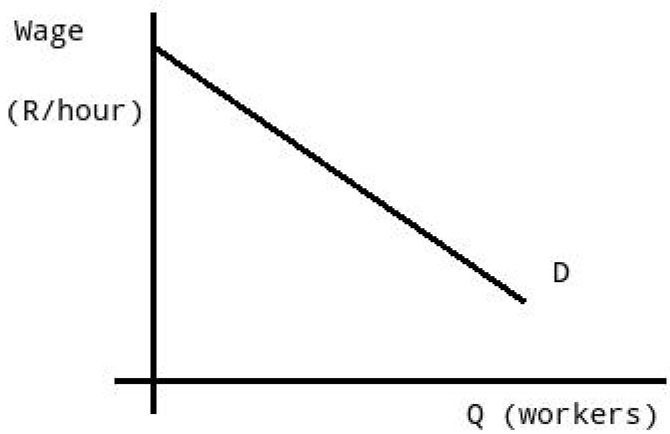


Figure 4.

Labor market equilibrium

In a perfectly operating labor markets the equilibrium wage rate (We) and employment levels (Qe) are determined by the interaction of market supply of and demand for labor as shown in the following sketch diagram. In reality, however, wages and employment levels are also influenced by such factors as government intervention (minimum wages etc.) as we have already seen. Trade unions also try to influence market wages and employment levels. So in real life labor markets, as in many other markets, are not perfectly ruled only by supply and demand for labor.

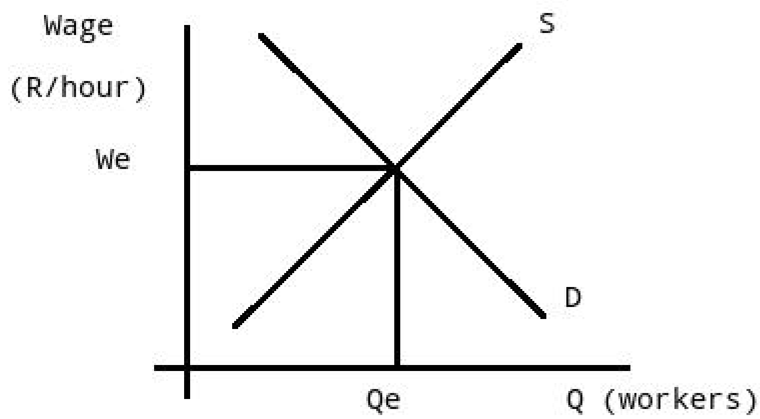


Figure 5.

The effect of trade unions on the labor market
We know that the real world is not perfect and that markets also are not perfect (that is, prices cannot always be relied upon to do a good rationing and allocation job according to the forces of market demand and market supply). For us studying the resource market (specifically the labor market) this means that wages (price of labor) and employment levels (quantity of labor) may be determined by forces other than just market supply and market demand. One example is the effect of trade unions.

Trade unions exist to protect the interests of workers since individual workers are generally at a disadvantage when negotiating with the employer. Trade unions try to influence working conditions and in particular the wage rates of their members. This is done by means of collective bargaining, mediation and arbitration. Strike action may be resorted to if mediation and arbitration fail. Trade unions generally take one of two possible forms: craft unions or industrial unions. We can use the supply and demand model to see how trade unions can influence the price of labor (wages) and employment levels.

Craft unions
Craft Unions consist of workers with a special set of skills. Examples of workers who would join such associations could include printers, plumbers,

electricians, pilots, jewelers etc. Examples in South Africa of these types of unions would include the South African Pilots Association, the South African Democratic Teacher's Union (SADTU) and the South African Football Player's Union. These types of unions can effectively control the supply of labor either by restricting membership, requiring stringent training/qualifications, controlling the duration of training etc. and so they can influence the equilibrium/market wage of labor. So instead of having a large supply of labor at S2 they limit the supply of labor to S1 (see diagram). This creates a shortage of that type of labor and so its price remains high at W1.

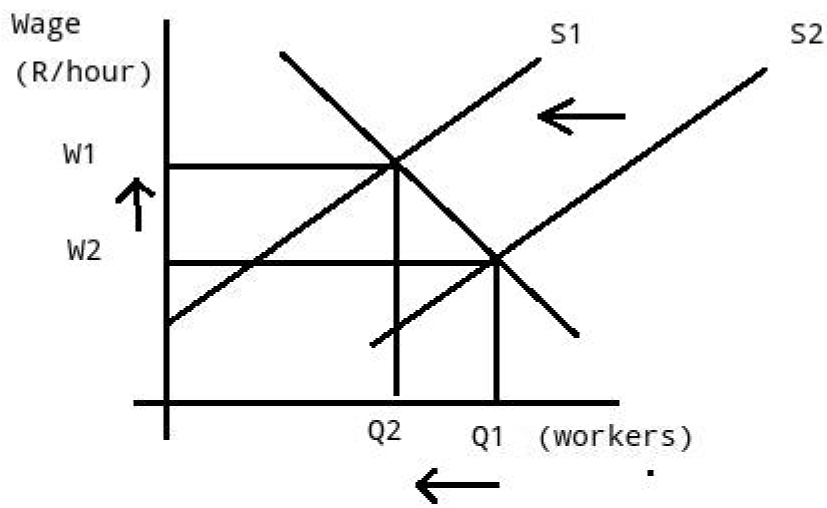


Figure 6.

Industrial unions (umbrella unions)

An industrial or "umbrella" union attempts to unite all workers (skilled and unskilled) in a particular industry into a single bargaining unit. Examples of such unions in South Africa would include the Congress of South African Trade Unions (COSATU),the National Union of Metal Workers of South Africa (NUMSA) and the Association of Mineworkers and Construction Union (AMCU). The goal of an industrial union is to achieve complete control over the labor supply thus forcing firms in the industry to bargain exclusively with it. However, because these tend to be large unions, it is

more difficult for them to control labor supply (unlike craft unions which are small enough to control their membership easily). Therefore industrial unions often resort to the threat of strike action to strengthen their position. Often just the threat of strike action may be enough to achieve their wage demands since employers are generally not keen to have production disrupted.

In the labor market diagram below, Wage1 is too low for union members. Due to threat of strike action (and the unwillingness of employers to face production losses) the union achieves an increase in wage to Wage2. This causes employers, over time, to reduce their quantity of labor demanded from e1 to e2. At the same time workers who were not previously interested in working at the old Wage1 are available to work at the new higher wage. The number of workers seeking employment increases to e3. Surplus labor = e2-e3 = unemployed. This is made up of e2-e1 (workers lost their jobs and not replaced) + e1-e3 (workers not available previously but now wanting to work). South African Trade unions have been criticized for protecting the interests of only their members and that they are not concerned with the country's serious unemployment problem. In addition strike action and associated violence (e.g. Marikana) have been blamed for causing bad relations between employers and employees and also for discouraging investment in South Africa.

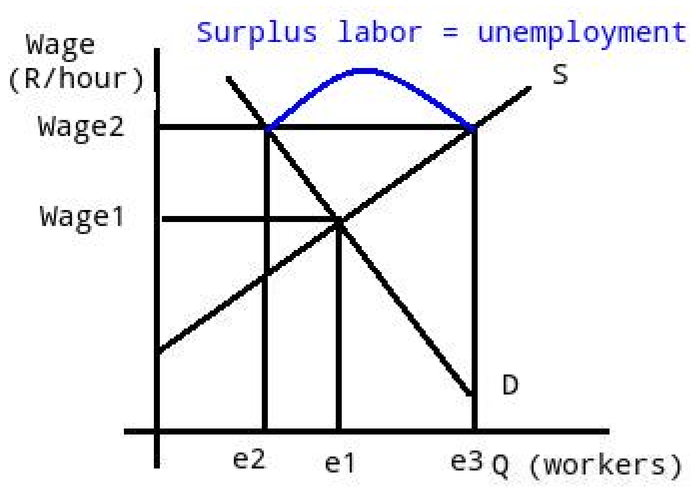


Figure 7.

A better way for employers and unions?
Perhaps a more constructive and peaceful way of increasing wage rates for its members is for the unions to enter into discussions/agreements with employers. Thus, if workers can demonstrate that they have added to the profitability of the firm then there is a stronger case for arguing for better wages. Profitability may be improved either by greater worker productivity and/or greater demand (and hence higher prices) for the firm's products. Both possibilities would increase the firm's demand for labor as shown below. Increasing labor demand will increase equilibrium wage and employment levels. Trade unions will be happy with higher wages and more members and employers will be happy because they will be able to pay the higher wages from their higher profits.

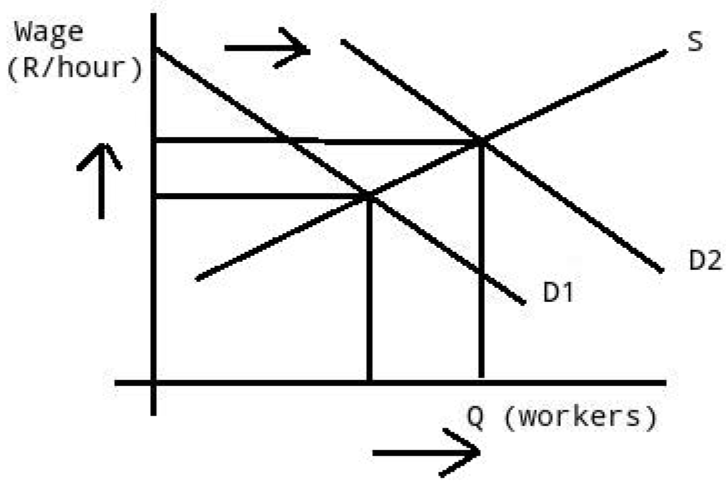


Figure 8.

A trade union may achieve its aim of increasing member wage rates by raising the productivity of its members (productivity agreements). This will increase the demand for labor and thus the wage rate and also employment levels as shown in the labor market diagram above. Unions could help

improve worker productivity by encouraging their members to attend training workshops facilitated by employers. Alternatively the firm's demand for labor can be increased by increasing the demand for goods and services produced by the firm (demand for labor is a derived demand). This will increase the equilibrium wage rate and also employment levels as shown in the diagram above. Unions could help promote or market the firm's produce by encouraging members and customers to buy from the firm and to buy South African produce for example ("buy local" campaigns).

Key Concepts and Summary

Price ceilings prevent a price from rising above a certain level. When a price ceiling is set below the equilibrium price, quantity demanded will exceed quantity supplied, and excess demand or shortages will result. Price floors prevent a price from falling below a certain level. When a price floor is set above the equilibrium price, quantity supplied will exceed quantity demanded, and excess supply or surpluses will result. Price floors and price ceilings often lead to unintended consequences.

Self-Check Questions

Exercise:

Problem:

What is the effect of a price ceiling on the quantity demanded of the product? What is the effect of a price ceiling on the quantity supplied? Why exactly does a price ceiling cause a shortage?

Solution:

A price ceiling (which is below the equilibrium price) will cause the quantity demanded to rise and the quantity supplied to fall. This is why a price ceiling creates a shortage.

Exercise:

Problem: Does a price ceiling change the equilibrium price?

Solution:

A price ceiling is just a legal restriction. Equilibrium is an economic condition. People may or may not obey the price ceiling, so the actual price may be at or above the price ceiling, but the price ceiling does not change the equilibrium price.

Exercise:

Problem:

What would be the impact of imposing a price floor below the equilibrium price?

Solution:

A price ceiling is a legal maximum price, but a price floor is a legal minimum price and, consequently, it would leave room for the price to rise to its equilibrium level. In other words, a price floor below equilibrium will not be binding and will have no effect.

Review Questions

Exercise:

Problem:

Does a price ceiling attempt to make a price higher or lower?

Exercise:

Problem:

How does a price ceiling set below the equilibrium level affect quantity demanded and quantity supplied?

Exercise:

Problem: Does a price floor attempt to make a price higher or lower?

Exercise:

Problem:

How does a price floor set above the equilibrium level affect quantity demanded and quantity supplied?

Critical Thinking Questions

Exercise:

Problem:

Most government policy decisions have winners and losers. What are the effects of raising the minimum wage? It is more complex than simply producers lose and workers gain. Who are the winners and who are the losers, and what exactly do they win and lose? To what extent does the policy change achieve its goals?

Exercise:

Problem:

Agricultural price supports result in governments holding large inventories of agricultural products. Why do you think the government cannot simply give the products away to poor people?

Exercise:

Problem:

Can you propose a policy that would make the market supply more rental housing units?

Exercise 11

Suppose government decides to set a price ceiling on bread so it can make sure that bread is affordable to the poor. The conditions of demand and

supply are given in the table below. What are the equilibrium price and equilibrium quantity before the price ceiling? What will the excess demand or the shortage (that is, quantity demanded minus quantity supplied) be if the price ceiling is set at R5/loaf? At R7/loaf? At R9/loaf?

P (R/loaf)	Loaves (millions/day)	Loaves (millions/day)
5	9000	5000
6	8500	5500
7	8000	6400
8	**7500**	**7500**
9	7000	9000
10	6500	11000
11	6000	15000

Table 2. Bread price schedule.

References

Mohamed, S.I. 2012. Rent rises: all you need to know. IOL news, 4 Sep. Available: http://www.iol.co.za/dailynews/2.1443/rent-rises-all-you-need-to-know-1375463 (Accessed: 23 Apr 2016)

Minimum wages in South Africa. 2016. Available: http://www.mywage.co.za/main/salary/minimum-wages (Accessed: 23 Apr 2016)

Nicolson, G. 2015. South Africa: where 12 million live in extreme poverty. Daily Maverick, 3 Feb. Available: http://www.dailymaverick.co.za/article/2015-02-03-south-africa-where-12-million-live-in-extreme-poverty/#.Vxupn3qwvIU (Accessed: 23 Apr 2016)

Glossary

price ceiling

a legal maximum price

price control

government laws to regulate prices instead of letting market forces determine prices

price floor

a legal minimum price

total surplus

see social surplus

Introduction to Elasticity
class="introduction"

Note:
Introduction to Elasticity
In this chapter, you will learn about:

- Price Elasticity of Demand and Price Elasticity of Supply
- Different elasticity values
- Elasticity and Pricing
- Elasticity in areas other than price

Note: The company Multichoice has been criticized for its 2016 DStv price increases (Koza: 2016). Multichoice maintains, however, that its prices are much the same as international TV packages. It has also pointed out that, while some of its prices have increased, others such as for the DStv Access package have remained unchanged while the price of DStv EasyView has actually fallen. What is behind the different pricing strategies for Multichoice's different packages? The theory of elasticity might help us to work out what the company is up to.

Figure 1. DSTV Logo (Credit: Wikimedia Commons, 2012)

	2015	1 April 2016
DStv Premium	R699	R759
DStv Extra	R425	R459
DStv Compact	R319	R345
DStv Select	R199	R219
DStv Family	R199	R219
DStv Access	R99	R99
DStv EasyView	R39	R29
M-Net analogue/CSN	R335	R369

Table 1. DST subscription rates, 2015 and 2016.

Anyone who has studied economics knows the law of demand: a higher price will lead to a lower quantity demanded. What you may not know is how much lower the quantity demanded will be. Similarly, the law of supply shows that a higher price will lead to a higher quantity supplied. The question is: how much higher? This chapter will explain how to answer these questions and why they are critically important in the real world.

To find answers to these questions, we need to understand the concept of elasticity. **Elasticity** is an economics concept that measures responsiveness of one variable to changes in another variable. Suppose you drop two items from the second-floor of a building. The first item is a tennis ball. The second item is a brick. Which will bounce higher? Obviously, the tennis ball. We would say that the tennis ball has greater elasticity.

Now consider an economic example. Cigarette taxes are an example of a “sin tax,” a tax on something that is bad for you, like alcohol. Cigarettes in South Africa in the 2016-2017 financial year are "sin" taxed at the rate of R13.24 per pack of 20 cigarettes. The sin taxes on alcohol are much less and also vary depending on the different types of alcohol that are sold. Beer, for example, is taxed at the rate of 135c/340ml can in 2016 while spirits are taxed at the rate of 394c/750ml. Why the different tax strategies for the different products? Is it because government is REALLY concerned about our health or could it also have to do with tax strategy and how best to collect the most tax? The theory of elasticity might help us understand this better.

Taxes on cigarettes serve two purposes: to raise tax revenue for government and to discourage consumption of cigarettes for obvious health reasons. However, if a higher cigarette tax discourages consumption by quite a lot, meaning a greatly reduced quantity of cigarettes is sold, then the cigarette tax on each pack will not raise much revenue for the government. Alternatively, a higher cigarette tax that does not discourage consumption by much will actually raise more tax revenue for the government. Thus, when the South African Revenue Service (SARS) tries to calculate the effects of altering its cigarette tax, it must analyze how much the tax affects

the quantity of cigarettes consumed. This issue reaches beyond governments and taxes. Every firm faces a similar issue. Every time a firm considers raising the price that it charges, it must consider how much a price increase will reduce the quantity demanded of what it sells. Conversely, when a firm puts its products on sale, it must expect (or hope) that the lower price will lead to a significantly higher quantity demanded.

References

Gadgenator. 2012. DStv’s Logo since October 2012. Image, 10 Oct. Available https://commons.wikimedia.org/wiki/File:DStv_Logo_2012.png. (Accessed: 24 Nov 2016)

Hawkins, C. 2016. DStv 2016 price hike announced. SA breaking news, 1 Mar. Available http://www.sabreakingnews.co.za/2016/03/01/dstv-2016-price-hike-announced/ (Accessed: 24 Apr 2016)

Koza, N. 2016. Complaints flood in over DSTV’s new price hike. Eyewitness news, 2 Mar. Available http://ewn.co.za/2016/03/02/Complaints-flood-in-over-DSTVs-new-price-hike (Accessed: 24 Apr 2016)

Sin taxes: prepare to pay more for cigarettes and booze. 2016. Mail and Guardian, 24 Feb. Available http://mg.co.za/article/2016-02-24-sin-taxes-prepare-to-pay-more-for-cigarettes-and-booze (Accessed: 24 Apr 2016)

Price Elasticity of Demand and Price Elasticity of Supply
By the end of this section, you will be able to:

- Calculate the price elasticity of demand
- Calculate the price elasticity of supply

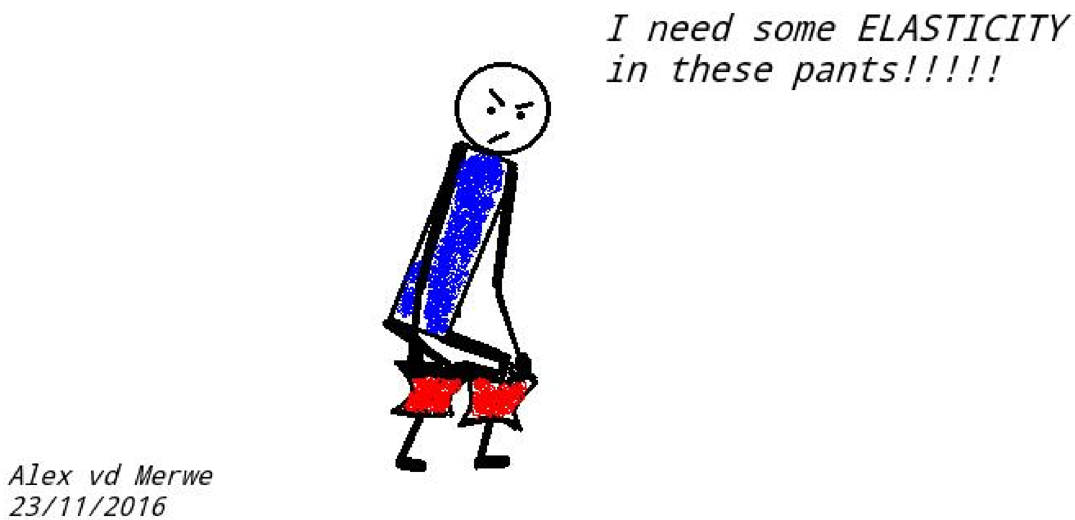


Figure 1. Econman demonstrates that the concept of "elasticity" in economics is not as frightening as it looks.

Both the demand and supply curve show the relationship between price and the number of units demanded or supplied. **Price elasticity** is the ratio between the percentage change in the quantity demanded (Qd) or supplied (Qs) and the corresponding percent change in price. The **price elasticity of demand** is the percentage change in the quantity *demanded* of a good or service divided by the percentage change in the price. The **price elasticity of supply** is the percentage change in quantity *supplied* divided by the percentage change in price.

Elasticities can be usefully divided into three broad categories: elastic, inelastic, and unitary. An **elastic demand** or **elastic supply** is one in which the elasticity is greater than one, indicating a high responsiveness to changes in price. Elasticities that are less than one indicate low responsiveness to price changes and correspond to **inelastic demand** or **inelastic supply**. **Unitary elasticities** indicate proportional responsiveness of either demand or supply, as summarized in Table 1.

If . . .	Then . . .	And It Is Called . . .
% change in quantity $>$ % change in price	$\frac{\text{% change in quantity}}{\text{% change in price}} > 1$	Elastic
% change in quantity $=$ % change in price	$\frac{\text{% change in quantity}}{\text{% change in price}} = 1$	Unitary
% change in quantity $<$ % change in price	$\frac{\text{% change in quantity}}{\text{% change in price}} < 1$	Inelastic

Elastic, Inelastic, and Unitary: Three Cases of Elasticity

Two different ways of calculating elasticity
Equation:

The point elasticity formula

Elasticity = (% change in quantity/% change in price) x 100

where % change is calculated as [(Q2 - Q1)/Q1] x 100 and % change in price is calculated as [(P2 - P1)/P1] x 100 with reference to the diagrams below in Figure 2.

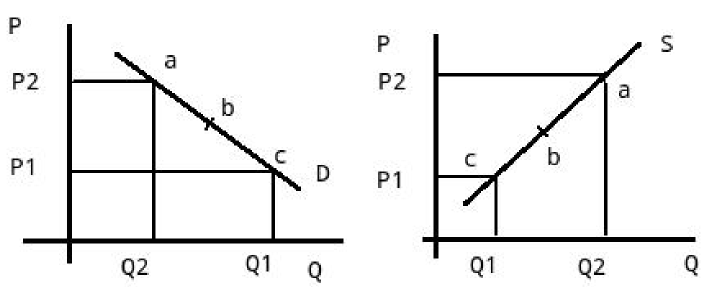


Figure 2.

So a price increase from P1 to p2 will result in a change in quantity demanded from Q1 to Q2 (whether demand or supply). The point elasticity value for this elasticity calculation (whether price elasticity of demand or price elasticity of supply) will be the value of the relevant elasticity coefficient at point c on the demand or supply curves (depending on which point elasticity is being computed). So this price elasticity formula gives the elasticity values at the points from which quantity changes in response to price changes.

The midpoint elasticity formula

The point elasticity formula will give slightly different elasticity values depending on the direction of change in price and quantity. So, for example, if price decreases from P2 tp P1, quantity will increase from Q1 to Q2. Even though the absolute sizes of the price and quantity changes are the same as for a price increase, the percent changes in these values will be different simply because their starting values (bases) are different and this will give different elasticity values. A way of getting around this problem is to calculate the elasticity at the midpoint between price and quantity changes on a demand or supply curve (if we are interested in price elasticity). So , instead of using simple percentage changes in quantity and price, economists use the average percent change in both quantity and price. This is called the Midpoint Method for Elasticity, and is represented in the following equations:

Equation:

$$\%\text{ change in quantity} = \frac{Q_2 - Q_1}{(Q_2 + Q_1)/2} \times 100$$

$$\%\text{ change in price} = \frac{P_2 - P_1}{(P_2 + P_1)/2} \times 100$$

These values are then inserted into the standard price elasticity formula: Elasticity = (% change in quantity demanded / % change in price) x 100.

So a price increase from P1 to P2 will yield the same elasticity value as for a price decrease from P2 to P1 (unlike point elasticity which will give slightly different values). The advantage of the Midpoint Method then is that one obtains the same elasticity value between two price points (at point b on the diagrams above) whether there is a price increase or decrease. This is because the formula uses the same base for both cases.

Calculating Price Elasticity of Demand

Let's calculate the elasticity between points A and B shown in Figure 3. We will first do this according to the point elasticity formula and then subsequently we will recalculate price elasticity of demand using the midpoint formula (just for demonstration purposes).

`Demand for Lukhanyo Mdingi Men's T-shirts`

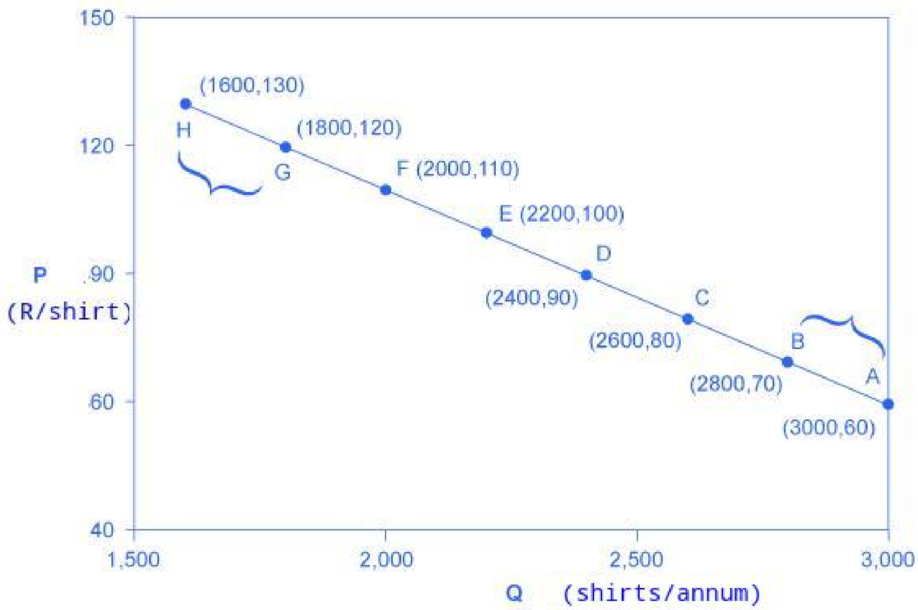


Figure 3.

Point elasticity computation at point A
The direction of change matters for point elasticity computations. So our point elasticity calculations in this chapter will be for price increases in each case. From point A to point B price increases by R10 per shirt and quantity demanded of shirts decreases by 200 shirts per annum from 3000 to 2800.

% change in price = [(P2-P1)/P1] x 100 = [(R70 - R60)/R60] x 100 = 16.7%

% change in quantity demanded = [(Q2-Q1)/Q1] x 100 = [(2800 - 3000)/3000] x 100 = -6.7%

Price elasticity of demand (Ep) = % change in quantity demanded/% change in price = -6.7%/167% = -0.4

The midpoint elasticity computation between points A and B
The direction of change does not matter with the midpoint formula. So whether price is increased or decreased it will yield the same elasticity value. Let's apply the formula to calculate the elasticity as price decreases from R70 at point B to R60 at point A:
Equation:

$$
\begin{aligned}
\text{\% change in quantity} &= \frac{3{,}000\text{–}2{,}800}{(3{,}000+2{,}800)/2} \times 100 \\
&= \frac{200}{2{,}900} \times 100 \\
&= 6.9 \\
\text{\% change in price} &= \frac{60\text{–}70}{(60+70)/2} \times 100 \\
&= \frac{\text{–}10}{65} \times 100 \\
&= \text{–}15.4 \\
\text{Price Elasticity of Demand} &= \frac{6.9\%}{\text{–}15.4\%} \\
&= -0.45
\end{aligned}
$$

Therefore, the elasticity of demand between these two points is $\frac{6.9\%}{-15.4\%}$ which is -0.45, an amount smaller than one, showing that the demand is inelastic in this interval. Note that this value is slightly higher than the point elasticity value at point A = -0.4 in absolute terms (ignoring the negative sign). Price elasticities of demand are *always* negative since price and quantity demanded always move in opposite directions (on the demand curve). It is generally accepted that price elasticities are given as positive values unless they have a value of zero. So mathematically, we take the absolute value (i.e. we ignore the negative sign) of the result. Always remember to interpret price elasticities as positive numbers!

For the midpoint elasticity value this means that, along the demand curve between point B and A, if the price falls by 1%, the quantity demanded will increase by 0.45% (by less than of the price change). For the point elasticity value an increase in price will result in a 0.4% decline in quantity demanded. In both cases, a change in the price will result in a smaller percentage change in the quantity demanded. Price elasticities of demand are negative numbers indicating that the demand curve is downward sloping, but are read as absolute values. The following Work It Out feature will walk you through calculating the price elasticity of demand.

Elasticity values do not have units of measurement

Another important point to remember about elasticity values is that they do not have units of measurement (e.g. Rands, percent etc.) since they are the product of percent changes. So elasticity values are simply coefficients (numbers or values) that measure by what fraction or multiple of price (in the case of price elasticity) quantity will change when price changes. Elasticity values are measured and interpreted with reference to the elasticity scale. Thus if elasticity exceeds 1 demand or supply is elastic, if elasticity is less than 1 demand or supply is inelastic and if elasticity = 1 demand or supply is unitary elastic.

Point elasticity computation at point B

It would be interesting to see how different the point elasticity value will be if we reverse the direction of the price change from B to A. In this case price falls from R70 per shirt to R60 per shirt and quantity demanded increases from 2800 to 3000 per annum. First we calculate the percent changes in price and quantities demanded.

% change in price = [(P2-P1)/P1] x 100 = [(R60 - R70)/R70] x 100 = -14.3%. P1 = R70 is now the starting price (the base) since it was stipulated that price falls from R70 to R60.

% change in quantity demanded = [(Q2-Q1)/Q1] x 100 = [(3000 - 2800)/2800] x 100 = 7.1%. Note that Q1 = 2800 is now the base since quantity demanded falls from 2800 to 3000 shirts per annum.

Price elasticity of demand (Ep) = % change in quantity demanded/% change in price = 7.1%/-14.3% = -0.5.

The midpoint elasticity value (midway between points A and B) should simply be the average of the point elasticity values at points A and B. So (-0.4+-0.5)/2 = -0.45. This is indeed the price elasticity of demand value computed using the midpoint formula!

Calculating the Price Elasticity of Supply

Assume that a letting agency rents out student accommodation for R650 per month and at that price 10,000 units are rented each year as shown in Figure 4. When the price increases to R700 per month, 13,000 units are supplied into the market. By what percentage does apartment supply increase? What is the price sensitivity? That is, how will quantity supplied of accommodation respond to change in price of accommodation?

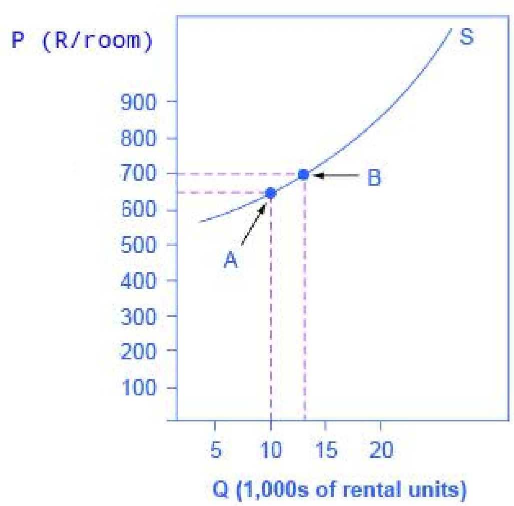


Figure 4: The price elasticity of supply is calculated as the percentage change in quantity divided by the percentage change in price.

Price elasticity of supply is computed in exactly the same way as price elasticity of demand with the only difference being that our focus is supply and not demand. So price elasticity of supply (Es) = % change in quantity supplied / % change in price. The resulting elasticity

coefficients are interpreted in exactly the same way as price elasticity of demand coefficients. So if Es exceeds 1 then supply is elastic, if Es is less than 1 supply is inelastic and if Es = 1 supply is unitary elastic.

Using the Point elasticity Method,

First we calculate % change in price = [(P2-P1)/P1] x 100 = [(R700 - R660)/R650] x 100 = 7.7%

Next we calculate % change in quantity demanded = [(Q2-Q1)/Q1] x 100 = [(13 000 - 10 000)/10 000] x 100 = 30%

Price elasticity of supply (Es) = % change in quantity supplied/% change in price = 30%/7.7% = 3.9. Supply is elastic.

Again, as with the elasticity of demand, the elasticity of supply is not followed by any units. It is a point worth repeating: Elasticity is a ratio of one percentage change to another percentage change—nothing more—and is read as an absolute value. It is worth noting, however, that elasticity of supply coefficients are generally positive due the law of supply (increase in price results in increase in quantity demanded). In this case, a 1% rise in price causes an increase in quantity supplied of 3.9%. The greater than one elasticity of supply means that the percentage change in quantity supplied will be greater than a one percent price change. If you're starting to wonder if the concept of slope fits into this calculation, read the following Clear It Up box.

Note:

Is the elasticity the slope?

It is a common mistake to confuse the slope of either the supply or demand curve with its elasticity. The slope is the rate of change in units along the curve, or the rise/run (change in y over the change in x). For example, each point on the demand for Lukhanyo Mdingi shirts indicates price intervals (changes) of R10 and the number of shirts demanded in 200 unit intervals. So the slope is –10/200 along the entire demand curve and does not change. The price elasticity, however, changes along the curve. Elasticity between points A and B is 0.45 while at point A it is 0.4 and at point B it is 0.5. So elasticity values are not constant along a straight line demand or supply curve whereas slope of a straight line is always constant (the same). Elasticity is the *percentage* change, which is a different calculation from the slope and has a different meaning.

When we are at the upper end of a demand curve (such as at points G and H on the demand for Lukhanyo Mdingi shirts), where price is high and the quantity demanded is low, a small change in the quantity demanded, even in, say, one unit, is pretty big in percentage terms. A change in price of, say, R10 per shirt, is going to be much less important in percentage terms than it would have been at the bottom of the demand curve. Likewise, at the bottom of the demand curve, a 200 unit change when the quantity demanded is high will be small as a percentage of a high quantity demanded.

So, at the top end of the demand curve (at relatively high prices), where we have a large percentage change in quantity demanded over a small percentage change in price, the elasticity values would be high, or demand would be relatively elastic (Ep greater than 1). Even with the same change in the price and the same change in the quantity demanded, at the bottom end of the demand curve the quantity is much higher, and the price is much lower, so the percentage change in quantity demanded is smaller and the percentage change in price is much higher. That means at the bottom of the curve we'd have a small numerator over a large denominator, so the elasticity measure would be much lower, or inelastic (Ep less than 1). And guess what? At the very middle point of every straight line demand curve the elasticity of demand value will always equal 1 (Ep = 1)!

As we move along the demand curve, the values for quantity and price go up or down, depending on which way we are moving, so the percentages for, say, a R1 difference in price or a one unit difference in quantity, will change as well, which means the ratios of those percentages will change. Which is why elasticity values change along the demand and supply curves.

Key Concepts and Summary

Price elasticity measures the responsiveness of the quantity demanded or supplied of a good to a change in its price. It is computed as the percentage change in quantity demanded (or supplied) divided by the percentage change in price. Elasticity can be described as elastic (or very responsive), unitary elastic, or inelastic (not very responsive). Elastic demand or supply curves indicate that quantity demanded or supplied respond to price changes in a greater than proportional manner. An inelastic demand or supply curve is one where a given percentage change in price will cause a smaller percentage change in quantity demanded or supplied. A unitary elasticity means that a given percentage change in price leads to an equal percentage change in quantity demanded or supplied.

Self-Check Questions

Exercise:

Problem:

From the data shown in Table 2 about demand for smart phones, calculate the point price elasticity of demand from: point B to point C (i.e. at point B), point D to point E (i.e. at point D), and point G to point H (at point G). Classify the elasticity at each point as elastic, inelastic, or unit elastic.

Points	P	Q
A	600	3,000
B	700	2,800
C	800	2,600
D	900	2,400
E	1000	2,200
F	1100	2,000
G	1200	1,800
H	1300	1,600

Solution:

B to C
Ep = (% change in Qd/% change in P) x 100 = -7.1%/14.3% = 0.5. Demand is inelastic.

D to E
Ep = (% change in Qd/% change in P) x 100 = -11.1%/8.3% = 0.74. Demand is still inelastic but not as inelastic as between B and C. It is becoming more elastic as price increases which is what we would expect.

G to H
Ep = (% change in Qd/% change in P) x 100 = -11.1%/8.3% = 1.3. Demand is elastic.

Review Questions

Exercise:

Problem: What is the formula for calculating elasticity?

Exercise:

Problem:

What is the price elasticity of demand? Can you explain it in your own words?

Exercise:

Problem: What is the price elasticity of supply? Can you explain it in your own words?

Critical Thinking Questions

Exercise:

Problem:

Air travel in business class has an estimated elasticity of demand of 0.40 less than air travel in economy class, with an estimated price elasticity of 0.62. Why do you think this is the case?

Exercise:

Problem:

What is the relationship between price elasticity and position on the demand curve? For example, as you move up the demand curve to higher prices and lower quantities, what happens to the measured elasticity? How would you explain that?

Problems

Exercise:

Problem:

The equation for a demand curve is P = 48 – 3Q. What is the elasticity in moving from a quantity of 5 to a quantity of 6?

Exercise:

Problem:

The equation for a demand curve is P = 2/Q. What is the elasticity of demand as price falls from 5 to 4? What is the elasticity of demand as the price falls from 9 to 8? Would you expect these answers to be the same?

Glossary

elastic demand
: when the elasticity of demand is greater than one, indicating a high responsiveness of quantity demanded or supplied to changes in price

elastic supply
: when the elasticity of either supply is greater than one, indicating a high responsiveness of quantity demanded or supplied to changes in price

elasticity

an economics concept that measures responsiveness of one variable to changes in another variable

inelastic demand

when the elasticity of demand is less than one, indicating that a 1 percent increase in price paid by the consumer leads to less than a 1 percent change in purchases (and vice versa); this indicates a low responsiveness by consumers to price changes

inelastic supply

when the elasticity of supply is less than one, indicating that a 1 percent increase in price paid to the firm will result in a less than 1 percent increase in production by the firm; this indicates a low responsiveness of the firm to price increases (and vice versa if prices drop)

price elasticity

the relationship between the percent change in price resulting in a corresponding percentage change in the quantity demanded or supplied

price elasticity of demand

percentage change in the quantity *demanded* of a good or service divided the percentage change in price

price elasticity of supply

percentage change in the quantity *supplied* divided by the percentage change in price

unitary elasticity

when the calculated elasticity is equal to one indicating that a change in the price of the good or service results in a proportional change in the quantity demanded or supplied

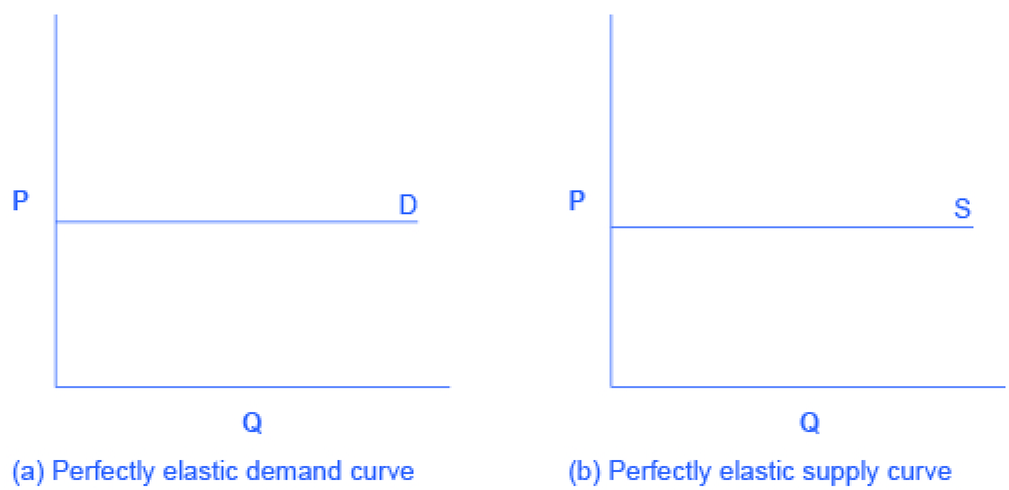
Polar Cases of Elasticity and Constant Elasticity

By the end of this section, you will be able to:

- Differentiate between infinite and zero elasticity
- Analyze graphs in order to classify elasticity as constant unitary, infinite, or zero

There are two extreme cases of elasticity: when elasticity equals zero and when it is infinite. A third case is that of constant unitary elasticity. We will describe each case. **Infinite elasticity** or **perfect elasticity** refers to the extreme case where either the quantity demanded (Qd) or supplied (Qs) changes by an infinite amount in response to any change in price at all. In both cases, the supply and the demand curve are horizontal as shown in Figure 1. While perfectly elastic supply curves are unrealistic, goods with readily available inputs and whose production can be easily expanded will feature highly elastic supply curves. Examples include pizza, bread, books, internet bandwidth, cellular telephone services and TV services. Similarly, perfectly elastic demand is an extreme example. But luxury goods, goods that take a large share of individuals' income, and goods with many substitutes are likely to have highly elastic demand curves. Examples of such goods are Caribbean cruises and sports vehicles. We will examine the factors that determine elasticity later in this module.

Infinite Elasticity

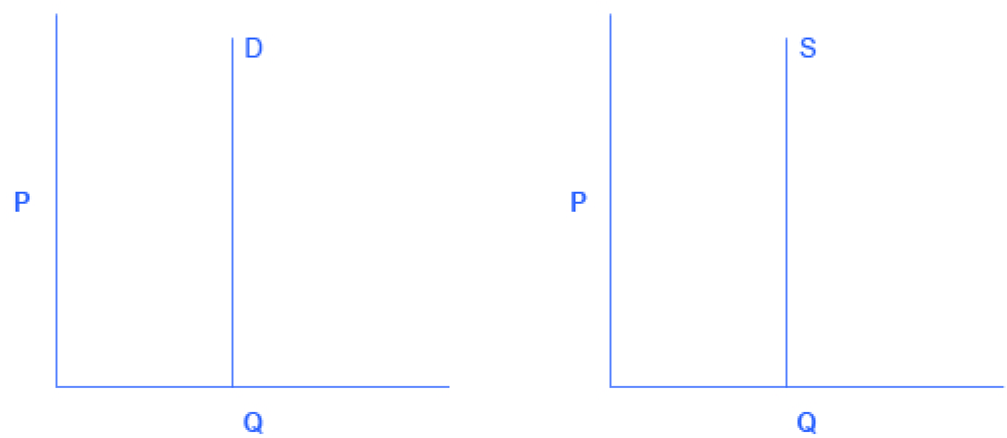


The horizontal lines show that an infinite quantity will be

demanded or supplied at a specific price. This illustrates the cases of a perfectly (or infinitely) elastic demand curve and supply curve. The quantity supplied or demanded is extremely responsive to price changes, moving from zero for prices close to P to infinite when price reaches P.

Zero elasticity or **perfect inelasticity**, as depicted in Figure 2 refers to the extreme case in which a percentage change in price, no matter how large, results in zero change in quantity. While a perfectly inelastic supply is an extreme example, goods with limited supply of inputs are likely to feature highly inelastic supply curves. Examples include diamond rings or housing in prime locations such as Montrose in Pietermaritzburg, Umhlanga in Durban and Clifton in Cape Town. Similarly, while perfectly inelastic demand is an extreme case, necessities with no close substitutes are likely to have highly inelastic demand curves. This is the case of life-saving medicines (ARVs, insulin for diabetes etc) and fuel.

Zero Elasticity

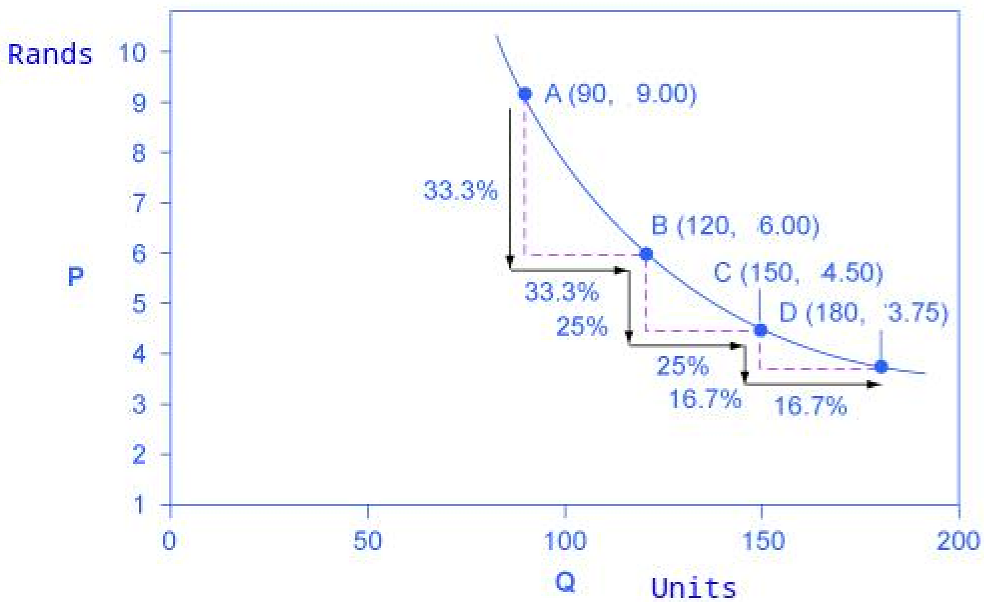


(a) Perfectly inelastic demand curve (b) Perfectly inelastic supply curve

The vertical supply curve and vertical demand curve show that there will be zero percentage change in quantity (a) demanded or (b) supplied, regardless of the price.

Constant unitary elasticity, in either a supply or demand curve, occurs when a price change of one percent results in a quantity change of one percent. Figure 3 shows a demand curve with constant unit elasticity. As we move down the demand curve from A to B, the price falls by 33% and quantity demanded rises by 33%; as you move from B to C, the price falls by 25% and the quantity demanded rises by 25%; as you move from C to D, the price falls by 16% and the quantity rises by 16%. Notice that in absolute value, the declines in price, as you step down the demand curve, are not identical. Instead, the price falls by R3 from A to B, by a smaller amount of R1.50 from B to C, and by a still smaller amount of R0.75 from C to D. As a result, a demand curve with constant unitary elasticity moves from a steeper slope on the left and a flatter slope on the right—and a curved shape overall.

A Constant Unitary Elasticity Demand Curve



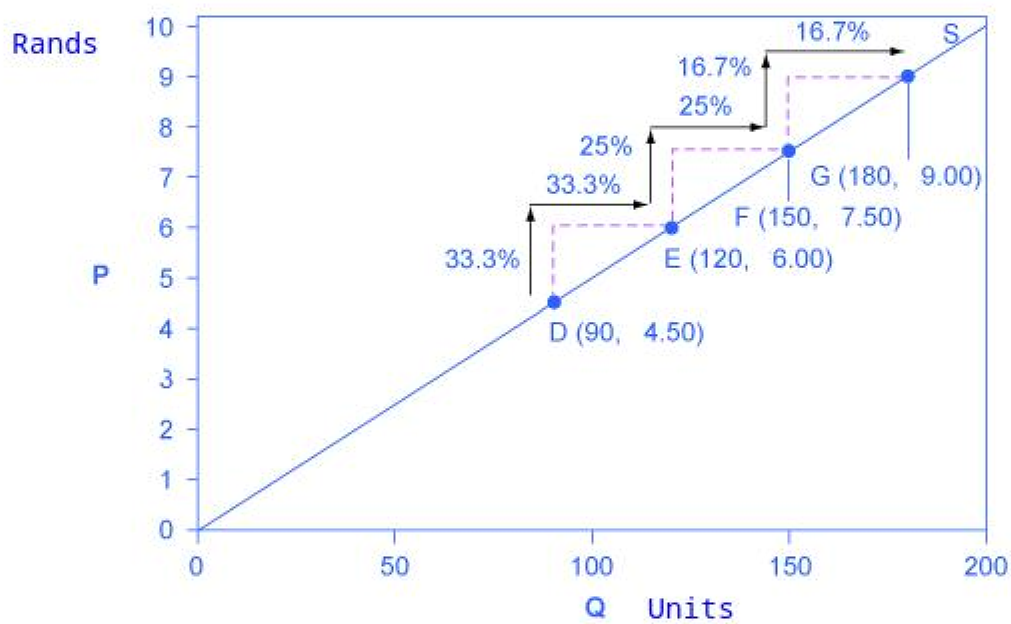
A demand curve with constant unitary elasticity will be a curved line. Notice how price and quantity demanded change by an identical amount in each step down the demand curve.

Unlike the demand curve with unitary elasticity, the supply curve with unitary elasticity is represented by a straight line. In moving up the supply

curve from left to right, each increase in quantity of 30, from 90 to 120 to 150 to 180, is equal in absolute value. However, in percentage value, the steps are decreasing, from 33.3% to 25% to 16.7%, because the original quantity points in each percentage calculation are getting larger and larger, which expands the denominator in the elasticity calculation.

Consider the price changes moving up the supply curve in Figure 4. From points D to E to F and to G on the supply curve, each step of R1.50 is the same in absolute value. However, if the price changes are measured in percentage change terms, they are also decreasing, from 33.3% to 25% to 16.7%, because the original price points in each percentage calculation are getting larger and larger in value. Along the constant unitary elasticity supply curve, the percentage quantity increases on the horizontal axis exactly match the percentage price increases on the vertical axis—so this supply curve has a constant unitary elasticity at all points.

A Constant Unitary Elasticity Supply Curve



A constant unitary elasticity supply curve is a straight line reaching up from the origin. Between each point, the percentage increase in quantity supplied is the same as the percentage increase in price.

Determinants of price elasticity of demand
We cannot say exactly what will make demand for a product more or less elastic. We can make some generalisations though. One is that the more **substitutes** a product has, the greater the elasticity of demand (simply because consumers have more choice). It can also depend on how broadly or narrowly the product is **defined**. The more broadly a product is defined, the smaller the elasticity of demand ("fruit" for example will have a smaller elasticity of demand than, say, "apples". Again, this is because fruit is associated with more choice).

Other things being equal, the higher the price of a good or service relative to peoples' **incomes** and therefore their budgets, the greater the elasticity of demand. So, for example, consumers are much more sensitive to an increase in the price of motor cars (high price relative to income) than to an increase in the price of candles (small price relative to income). Also, the more **durable** (long lasting) a product is, the higher the elasticity of demand simply because consumers can put off the decision to purchase a replacement product should they need to (e.g. a fridge that does not work very well can still do a job but milk or bread that is old or bad cannot be kept).

The demand for **necessities** (bread, cooking oil, medicines, fuel, electricity) tends to be price inelastic while for **luxuries** (designer/boutique clothing and shoes, overseas holidays, luxury cars etc.) it tends to be price elastic. So when price of necessities increases consumers have little choice but to continue purchasing them. However, they do not have to purchase luxury items. Finally, product demand is more elastic the longer the **time period** under consideration. And once again, this is because consumers have greater choice when they have more time to shop around for alternatives (substitutes) if price increases.

Determinants of price elasticity of supply
The amount that costs rise as output rises will influence the elasticity of supply. So, the smaller the additional cost of producing additional output, the more firms will be encouraged to to produce for a given price increase and the more elastic supply will be. Supply will be more elastic if firms have plenty of spare capacity, if they can easily get extra supplies of raw

materials, if they can easily switch away from producing alternative products and if they can hold their costs of production down. If these conditions hold then costs will be little affected by an increase in output and so firms can more easily and quickly respond to price increases: supply will be relatively elastic. And of course the more **time** firms have to respond to changes in market conditions (e.g. increase in price of their products), the more easily they can increase output: the more elastic supply will be. Short run (term) supply is relatively inelastic compared with long run supply.

Figure 5.

Key Concepts and Summary

Infinite or perfect elasticity refers to the extreme case where either the quantity demanded or supplied changes by an infinite amount in response to any change in price at all. Zero elasticity refers to the extreme case in which

a percentage change in price, no matter how large, results in zero change in quantity. Constant unitary elasticity in either a supply or demand curve refers to a situation where a price change of one percent results in a quantity change of one percent.

Review Questions

Exercise 1
Identify and briefly discuss three determinants of price elasticity of demand.

Exercise 2
Describe the general appearance of a demand curve with infinite elasticity.

Exercise 3
Describe the general appearance of a demand curve with zero elasticity.

Exercise 4
Briefly discuss one determinant of the elasticity of supply.

Critical Thinking Question

Exercise 5
Can you think of an industry (or product) with near infinite elasticity of supply in the short term? That is, what is an industry that could increase Qs almost without limit in response to an increase in the price?

Problems

Exercise 6
The supply of paintings by Leonardo Da Vinci, who painted the Mona Lisa and The Last Supper and died in 1519, is highly inelastic. Sketch a supply and demand diagram, paying attention to the appropriate elasticities, to illustrate that demand for these paintings will determine the price.

Exercise 7
Say that a certain stadium for professional football has 70,000 seats. What is the shape of the supply curve for tickets to football games at that

stadium? Explain.

Exercise 8

When someone's kidneys fail, the person needs to have medical treatment with a dialysis machine (unless or until they receive a kidney transplant) or they will die. Sketch a supply and demand diagram, paying attention to the appropriate elasticities, to illustrate that the supply of such dialysis machines will primarily determine the price.

Glossary

constant unitary elasticity
: when a given percent price change in price leads to an equal percentage change in quantity demanded or supplied

infinite elasticity
: the extremely elastic situation of demand or supply where quantity changes by an infinite amount in response to any change in price; horizontal in appearance

perfect elasticity
: see infinite elasticity

zero inelasticity
: the highly inelastic case of demand or supply in which a percentage change in price, no matter how large, results in zero change in the quantity; vertical in appearance

perfect inelasticity
: see zero elasticity

Elasticity and Pricing

By the end of this section, you will be able to:

- Analyze how price elasticities impact revenue
- Evaluate how elasticity can cause shifts in demand and supply
- Predict how the long-run and short-run impacts of elasticity affect equilibrium
- Explain how the elasticity of demand and supply determine the incidence of a tax on buyers and sellers

Studying elasticities is useful for a number of reasons, pricing (how firms price their products) being most important. Let’s explore how elasticity relates to revenue and pricing, both in the long run and short run. But first, let’s look at the elasticities of some common goods and services.

Table 1 shows a selection of demand elasticities for different goods and services drawn from a variety of different studies by economists, listed in order of increasing elasticity.

Goods and Services	**Elasticity of Price**
Housing	0.12
Domestic air travel (economy class)	0.12
Kombi transport (rush hour)	0.15
Electricity	0.20
Taxi cabs	0.22

Goods and Services	**Elasticity of Price**
Petrol	0.35
International air travel (first class)	0.40
Wine	0.55
Beef	0.59
International air travel (business class)	0.62
Kitchen and household appliances	0.63
Dstv (basic version)	0.69
Chicken	0.64
Soft drinks	0.70
Beer	0.80
New vehicle	0.87
Rail travel (off-peak)	1.00
Computer	1.44
Dstv (mid-range version)	1.51
Dstv (premium version)	1.77
Restaurant meals	2.27

Some Selected Elasticities of Demand

Note that necessities such as housing and electricity are inelastic (these could be considered essential products, remember the determinants of elasticity of demand), while items that are not necessities such as restaurant meals (luxury items) are more price-sensitive. If the price of the restaurant meal increases by 10%, the quantity demanded will decrease by 22.7%. A 10% increase in the price of housing will cause a slight decrease of 1.2% in the quantity of housing demanded.

Does Raising Price Bring in More Revenue?

Imagine that a band on tour is playing in an indoor arena with 15,000 seats. To keep this example simple, assume that the band keeps all the money from ticket sales. Assume further that the band pays the costs for its appearance, but that these costs, like travel, setting up the stage, and so on, are the same regardless of how many people are in the audience. Finally, assume that all the tickets have the same price. (The same insights apply if ticket prices are more expensive for some seats than for others, but the calculations become more complicated.) The band knows that it faces a downward-sloping demand curve; that is, if the band raises the price of tickets, it will sell fewer tickets. How should the band set the price for tickets to bring in the most total revenue, which in this example, because costs are fixed, will also mean the highest profits for the band? Should the band sell more tickets at a lower price or fewer tickets at a higher price?

The key concept in thinking about collecting the most revenue is the price elasticity of demand. Total revenue is price times the quantity of tickets sold. Imagine that the band starts off thinking about a certain price, which will result in the sale of a certain quantity of tickets. The three possibilities are laid out in Table 2. If demand is elastic at that price level, then the band should cut the price, because the percentage drop in price will result in an even larger percentage increase in the quantity sold—thus raising total revenue. However, if demand is inelastic at that original quantity level, then the band should raise the price of tickets, because a certain percentage increase in price will result in a smaller percentage decrease in the quantity sold—and total revenue will rise. If demand has a unitary elasticity at that quantity, then a moderate percentage change in the price will be offset by an

equal percentage change in quantity—so the band will earn the same revenue whether it (moderately) increases or decreases the price of tickets.

If Demand Is . . .	**Then . . .**	**Therefore . . .**
Elastic	$\%\ \text{change in Qd} > \%\ \text{change in P}$	A given % rise in P will be more than offset by a larger % fall in Q so that total revenue (P × Q) falls.
Unitary	$\%\ \text{change in Qd} = \%\ \text{change in P}$	A given % rise in P will be exactly offset by an equal % fall in Q so that total revenue (P × Q) is unchanged.

If Demand Is . . .	**Then . . .**	**Therefore . . .**
Inelastic	$\%$ change in Qd $< \%$ change in P	A given % rise in P will cause a smaller % fall in Q so that total revenue (P × Q) rises.

Will the Band Earn More Revenue by Changing Ticket Prices?

What if the band keeps cutting price, because demand is elastic, until it reaches a level where all 15,000 seats in the available arena are sold? If demand remains elastic at that quantity, the band might try to move to a bigger arena, so that it could cut ticket prices further and see a larger percentage increase in the quantity of tickets sold. Of course, if the 15,000-seat arena is all that is available or if a larger arena would add substantially to costs, then this option may not work.

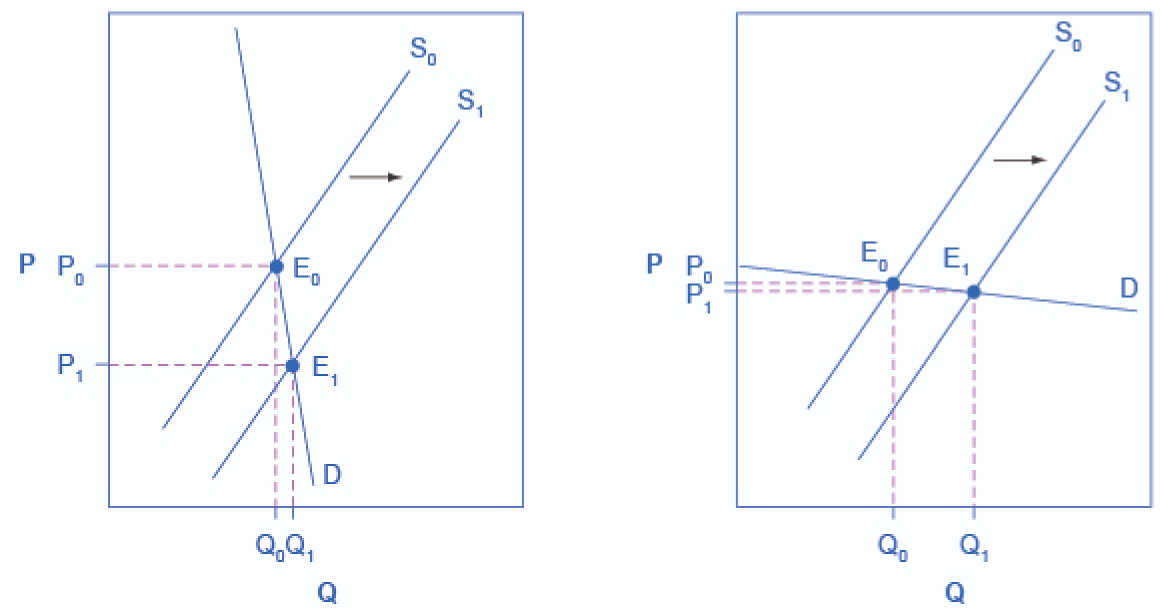
Conversely, a few bands are so famous, or have such fanatical followings, that demand for tickets may be inelastic right up to the point where the arena is full. These bands can, if they wish, keep raising the price of tickets. Ironically, some of the most popular bands could make more revenue by setting prices so high that the arena is not filled—but those who buy the tickets would have to pay very high prices. However, bands sometimes choose to sell tickets for less than the absolute maximum they might be able to charge, often in the hope that fans will feel happier and spend more on recordings, T-shirts, and other band memorabilia.

Can Costs Be Passed on to Consumers?

Most businesses face a day-to-day struggle to figure out ways to produce at a lower cost, as one pathway to their goal of earning higher profits. However, in some cases, the price of a key input over which the firm has no control may rise. For example, many chemical companies use petroleum as a key input, but they have no control over the world market price for crude oil. Coffee shops use coffee as a key input, but they have no control over the world market price of coffee. If the cost of a key input rises, can the firm pass those higher costs along to consumers in the form of higher prices? Conversely, if new and less expensive ways of producing are invented, can the firm keep the benefits in the form of higher profits, or will the market pressure them to pass the gains along to consumers in the form of lower prices? The price elasticity of demand plays a key role in answering these questions.

Imagine that as a consumer of legal pharmaceutical products, you read a newspaper story that a technological breakthrough in the production of aspirin has occurred, so that every aspirin factory can now make aspirin more cheaply than it did before. What does this discovery mean to you? Figure 1 illustrates two possibilities. In Figure 1 (a), the demand curve is drawn as highly inelastic. In this case, a technological breakthrough that shifts supply to the right, from S_0 to S_1, so that the equilibrium shifts from E_0 to E_1, creates a substantially lower price for the product with relatively little impact on the quantity sold. In Figure 1 (b), the demand curve is drawn as highly elastic. In this case, the technological breakthrough leads to a much greater quantity being sold in the market at very close to the original price. Consumers benefit more, in general, when the demand curve is more inelastic because the shift in the supply results in a much lower price for consumers.

Passing along Cost Savings to Consumers



(a) Cost-saving with inelastic demand

(b) Cost-saving with elastic demand

Cost-saving gains cause supply to shift out to the right from S_0 to S_1; that is, at any given price, firms will be willing to supply a greater quantity. If demand is inelastic, as in (a), the result of this cost-saving technological improvement will be substantially lower prices. If demand is elastic, as in (b), the result will be only slightly lower prices. Consumers benefit in either case, from a greater quantity at a lower price, but the benefit is greater when demand is inelastic, as in (a).

Producers of aspirin may find themselves in a difficult position here. The situation shown in Figure 1 (a), with extremely inelastic demand, means that a new invention may cause the price to drop dramatically while quantity changes little. As a result, the new production technology can lead to a drop in the revenue that firms earn from sales of aspirin. However, if strong competition exists between producers of aspirin, each producer may have little choice but to search for and implement any breakthrough that allows it to reduce production costs. After all, if one firm decides not to implement such a cost-saving technology, it can be driven out of business by other firms that do.

Since demand for food is generally inelastic, farmers may often face the situation in Figure 1 (a). That is, a surge in production leads to a severe drop in price that can actually decrease the total revenue received by farmers. Conversely, poor weather or other conditions that cause a terrible year for farm production can sharply raise prices so that the total revenue received increases. The Clear It Up box discusses how these issues relate to coffee.

Note:

How do coffee prices fluctuate?

Coffee is an international crop. The top five coffee-exporting nations are Brazil, Vietnam, Colombia, Indonesia, and Ethiopia. In these nations and others, 20 million families depend on selling coffee beans as their main source of income. These families are exposed to enormous risk, because the world price of coffee bounces up and down. For example, in 1993, the world price of coffee (priced in US Dollars) was about $1.10 per Kg; in 1995 it was four times as high, at $4.40 per Kg. By 1997 it had fallen by half to $2.20 per Kg. In 1998 it leaped back up to $4.40 per Kg. By 2001 it had fallen back to $1.10/Kg; by early 2011 it went back up to about $5.82/Kg. By the end of 2012, the price had fallen back to about $2.88 per Kg.

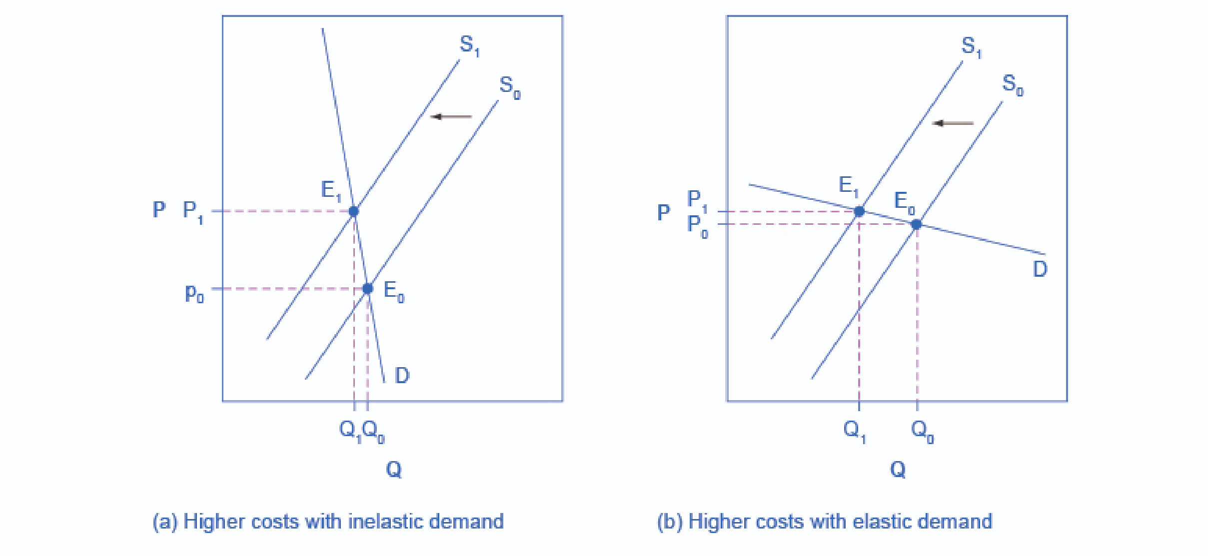
The reason for these price bounces lies in a combination of inelastic demand and shifts in supply. The elasticity of coffee demand is only about 0.3; that is, a 10% rise in the price of coffee leads to a decline of about 3% in the quantity of coffee consumed. When a major frost hit the Brazilian coffee crop in 1994, coffee supply shifted to the left with an inelastic demand curve, leading to much higher prices. Conversely, when Vietnam entered the world coffee market as a major producer in the late 1990s, the supply curve shifted out to the right. With a highly inelastic demand curve, coffee prices fell dramatically. This situation is shown in Figure 1 (a).

Elasticity also reveals whether firms can pass higher costs that they incur on to consumers. Addictive substances tend to fall into this category. For example, the demand for cigarettes is relatively inelastic among regular

smokers who are somewhat addicted; economic research suggests that increasing the price of cigarettes by 10% leads to about a 3% reduction in the quantity of cigarettes smoked by adults, so the elasticity of demand for cigarettes is 0.3. If society (through government) increases taxes on companies that make cigarettes, the result will be, as in Figure 2 (a), that the supply curve shifts from S_0 to S_1. However, as the equilibrium moves from E_0 to E_1, these taxes are mainly passed along to consumers in the form of higher prices. These higher taxes on cigarettes (excise taxes) will raise tax revenue for the government, but they will not much affect the quantity of smoking.

If the goal is to reduce the quantity of cigarettes demanded, it must be achieved by shifting this inelastic demand back to the left, perhaps with public programs to discourage the use of cigarettes or to help people to quit. For example, anti-smoking advertising campaigns have shown some ability to reduce smoking. However, if demand for cigarettes was more elastic, as in Figure 2 (b), then an increase in taxes that shifts supply from S_0 to S_1 and equilibrium from E_0 to E_1 would reduce the quantity of cigarettes smoked substantially. Youth smoking seems to be more elastic than adult smoking—that is, the quantity of youth smoking will fall by a greater percentage than the quantity of adult smoking in response to a given percentage increase in price.

Passing along Higher Costs to Consumers



Higher costs, like a higher tax on cigarette companies for the example

given in the text, lead supply to shift to the left. This shift is identical in (a) and (b). However, in (a), where demand is inelastic, the cost increase can largely be passed along to consumers in the form of higher prices, without much of a decline in equilibrium quantity. In (b), demand is elastic, so the shift in supply results primarily in a lower equilibrium quantity. Consumers suffer in either case, but in (a), they suffer from paying a higher price for the same quantity, while in (b), they suffer from buying a lower quantity (and presumably needing to shift their consumption elsewhere).

Elasticity and Tax Incidence (tax burden)

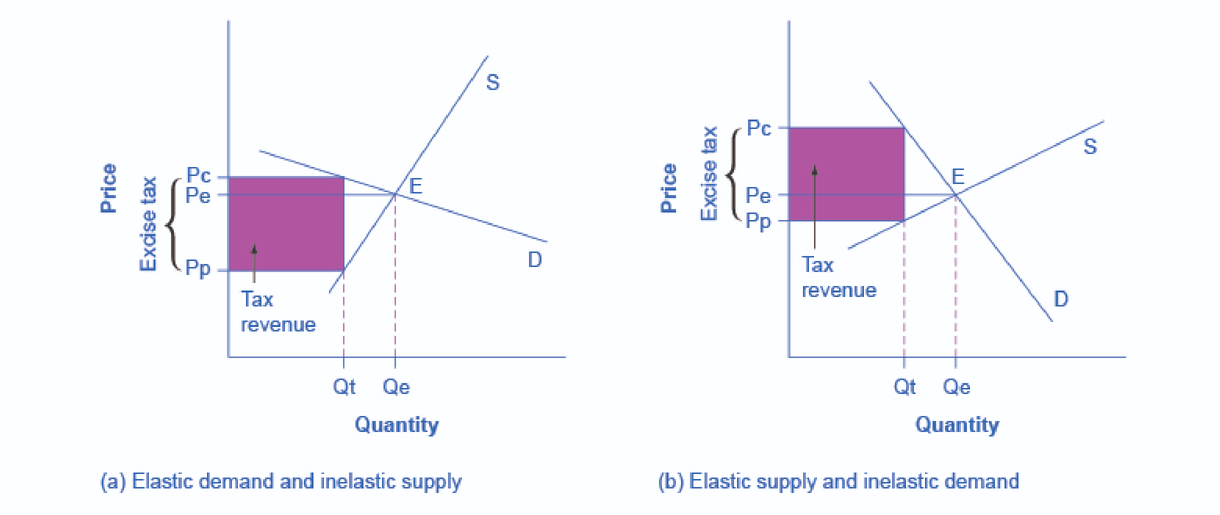
The example of cigarette taxes showed that because demand is inelastic, taxes are not effective at reducing the equilibrium quantity of smoking, and they are mainly passed along to consumers in the form of higher prices. The analysis, or manner, of how the burden of a tax is divided between consumers and producers is called **tax incidence**. Typically, the incidence, or burden, of a tax falls both on the consumers and producers of the taxed good. But if one wants to predict which group will bear most of the burden, all one needs to do is examine the elasticity of demand and supply. In the tobacco example, the tax burden falls on the most inelastic side of the market.

If demand is more inelastic than supply, consumers bear most of the tax burden, and if supply is more inelastic than demand, sellers bear most of the tax burden.

The logic of this is simple. When the demand is inelastic, consumers are not very responsive to price changes, and the quantity demanded remains relatively constant when the tax is introduced. In the case of smoking, the demand is inelastic because consumers are addicted to the product. The government can then pass the tax burden along to consumers in the form of higher prices, without much of a decline in the equilibrium quantity.

Similarly, when a tax is introduced in a market with an inelastic supply, such as, for example, beachfront hotels (or vegetable or dairy farmers), and sellers have no alternative than to accept lower prices for their business or produce, taxes do not greatly affect the equilibrium quantity. The tax burden is now passed on to the sellers. If the supply was elastic and sellers had the possibility of reorganizing their businesses to avoid supplying the taxed good, the tax burden on the sellers would be much smaller. The tax would result in a much lower quantity sold instead of lower prices received. Figure 3 illustrates this relationship between the tax incidence and elasticity of demand and supply.

Elasticity and Tax Incidence



An excise tax introduces a wedge between the price paid by consumers (Pc) and the price received by producers (Pp). (a) When the demand is more elastic than supply, the tax incidence on consumers Pc – Pe is lower than the tax incidence on producers Pe – Pp. (b) When the supply is more elastic than demand, the tax incidence on consumers Pc – Pe is larger than the tax incidence on producers Pe – Pp. The more elastic the demand and supply curves are, the lower the tax revenue.

In Figure 3 (a), the supply is inelastic and the demand is elastic, such as in the example of beachfront hotels (or dairy or vegetable farmers). While consumers may have other vacation choices, hoteliers can’t easily move their businesses. In the same way, farmers are stuck with perishable produce

that will not keep if they do not sell it soon. By introducing a tax, the government essentially creates a wedge between the price paid by consumers Pc and the price received by producers Pp. In other words, of the total price paid by consumers, part is retained by the sellers and part is paid to the government in the form of a tax. The distance between Pc and Pp is the tax rate. The new market price is Pc, but sellers receive only Pp per unit sold, as they pay Pc-Pp to the government. Since a tax can be viewed as raising the costs of production, this could also be represented by a leftward shift of the supply curve, where the new supply curve would intercept the demand at the new quantity Qt. For simplicity, Figure 3 omits the shift in the supply curve.

The tax revenue is given by the shaded area, which is obtained by multiplying the tax per unit by the total quantity sold Qt. The tax incidence (burden of the tax) on the consumers is given by the difference between the price paid Pc and the initial equilibrium price Pe. The tax incidence on the sellers is given by the difference between the initial equilibrium price Pe and the price they receive after the tax is introduced Pp. In Figure 3 (a), the tax burden falls disproportionately on the sellers, and a larger proportion of the tax revenue (the shaded area) is due to the resulting lower price received by the sellers than by the resulting higher prices paid by the buyers. The example of the tobacco excise tax could be described by Figure 3 (b) where the supply is more elastic than demand. The tax incidence now falls disproportionately on consumers, as shown by the large difference between the price they pay, Pc, and the initial equilibrium price, Pe. Sellers receive a lower price than before the tax, but this difference is much smaller than the change in consumers' price. From this analysis one can also predict whether a tax is likely to create a large revenue for government or not. The more elastic the demand curve, the easier it is for consumers to reduce quantity instead of paying higher prices. The more elastic the supply curve, the easier it is for sellers to reduce the quantity sold, instead of taking lower prices. In a market where both the demand and supply are very elastic, the imposition of an excise tax generates low tax revenue for government.

Excise taxes tend to be thought to hurt mainly the specific industries they target. For example, the annual special tax increases on alcohol and tobacco products announced each year by Minister of Finance and the National

Treasury medical device excise tax, in effect since 2013, has been controversial for it can delay industry profitability and even promote illicit trade (Miti, 2014). But ultimately, whether the tax burden falls mostly on the industry concerned or on consumers depends simply on the elasticity of demand and supply.

Long-Run vs. Short-Run Impact

Elasticities are often lower in the short run than in the long run. On the demand side of the market, it can sometimes be difficult to change Qd in the short run, but easier in the long run (recall that time period is a determinant of elasticity). The longer the time period, the more opportunity there is to find substitutes. Consumption of electricity or petrol are clear examples. In the short run, it is not easy for a person to make substantial changes in the electricity or petrol consumption if their prices increase. Maybe you can share lifts to school or work sometimes if the price of petrol increases or perhaps you can turn off the hot water geyser in the house when the electricity tariff increases. But that is about all you can do right now (i.e. in the short run). You still need to drive to school or work somehow and at some point you will have to switch the geyser back on. However, in the long-run you could purchase a car that is more fuel efficient, choose a job that is closer to where you live, buy more energy-efficient home appliances, or install a solar geyser in your home. As a result, the elasticity of demand for petrol and electricity is somewhat inelastic in the short run, but much more elastic in the long run.

On the supply side of markets, producers of goods and services typically find it easier to expand production in the long term of several years rather than in the short run of a few months. After all, in the short run it can be costly or difficult to build a new factory, hire many new workers, or open new stores. But over a few years, all of these are possible.

Indeed, in most markets for goods and services, prices bounce up and down more than quantities in the short run, but quantities often move more than prices in the long run. The underlying reason for this pattern is that supply and demand are often inelastic in the short run, so that shifts in either demand or supply can cause a relatively greater change in prices. But since

supply and demand are more elastic in the long run, the long-run movements in prices are not so severe, while quantity adjusts more easily in the long run.

Key Concepts and Summary

Figure 4. Price and income elasticity are useful measures especially in the field of marketing.

In the market for goods and services, quantity supplied and quantity demanded are often relatively slow to react to changes in price in the short run, but react more substantially in the long run. As a result, demand and

supply often (but not always) tend to be relatively inelastic in the short run and relatively elastic in the long run. The tax incidence depends on the relative price elasticity of supply and demand. When supply is more elastic than demand, buyers bear most of the tax burden, and when demand is more elastic than supply, producers bear most of the cost of the tax. Tax revenue is larger the more inelastic the demand and supply are.

Self-Check Questions

Exercise:

Problem:

The government decides to require that car manufacturers install new anti-pollution equipment that costs R31,000 per car. Under what conditions can carmakers pass almost all of this cost along to car buyers? Under what conditions can carmakers pass very little of this cost along to car buyers?

Solution:

Carmakers can pass this cost along to consumers if the demand for these cars is inelastic. If the demand for these cars is elastic, then the manufacturer must pay for the equipment.

Exercise:

Problem:

Suppose you are in charge of sales at a pharmaceutical company, and your firm has a new medication that causes bald men to grow hair. Assume that the company wants to earn as much revenue as possible from this drug. If the elasticity of demand for your company’s product at the current price is 1.4, would you advise the company to raise the price, lower the price, or to keep the price the same? What if the elasticity were 0.6? What if it were 1? Explain your answer.

Solution:

If the elasticity is 1.4 at current prices, you would advise the company to lower its price on the product, since a decrease in price will be offset by the increase in the amount of the drug sold. If the elasticity were 0.6, then you would advise the company to increase its price. Increases in price will offset the decrease in number of units sold, but increase your total revenue. If elasticity is 1, the total revenue is already maximized, and you would advise that the company maintain its current price level.

Review Questions

Exercise:

Problem:

If demand is elastic, will shifts in supply have a larger effect on equilibrium quantity or on price?

Exercise:

Problem:

If demand is inelastic, will shifts in supply have a larger effect on equilibrium price or on quantity?

Exercise:

Problem:

If supply is elastic, will shifts in demand have a larger effect on equilibrium quantity or on price?

Exercise:

Problem:

If supply is inelastic, will shifts in demand have a larger effect on equilibrium price or on quantity?

Exercise:

Problem:

Would you usually expect elasticity of demand or supply to be higher in the short run or in the long run? Why?

Exercise:

Problem:

Under which circumstances does the tax burden fall entirely on consumers?

Critical Thinking Questions

Exercise:

Problem:

Would you expect supply to play a more significant role in determining the price of a basic necessity like food or a luxury like perfume? Explain. *Hint*: Think about how the price elasticity of demand will differ between necessities and luxuries.

Exercise:

Problem:

A city has built a bridge over a river and it decides to charge a toll to everyone who crosses. For one year, the city charges a variety of different tolls and records information on how many drivers cross the bridge. The city thus gathers information about elasticity of demand. If the city wishes to raise as much revenue as possible from the tolls, where will the city decide to charge a toll: in the inelastic portion of the demand curve, the elastic portion of the demand curve, or the unit elastic portion? Explain.

Exercise:

Problem:

In a market where the supply curve is perfectly inelastic, how does an excise tax affect the price paid by consumers and the quantity bought and sold?

Problems

Exercise:

Problem:

Assume that the supply of low-skilled workers is fairly elastic, but the employers' demand for such workers is fairly inelastic. If the policy goal is to expand employment for low-skilled workers, is it better to focus on policy tools to shift the supply of unskilled labor or on tools to shift the demand for unskilled labor? What if the policy goal is to raise wages for this group? Explain your answers with supply and demand diagrams.

References

Miti, S., 2014. Tobacco tax hikes 'boost illegal traders. DispatchLive. Available: http://www.dispatchlive.co.za/news/tobacco-tax-hikes-boost-illegal-traders/ (Accessed: 11 April 2016)

Glossary

tax incidence
: manner in which the tax burden is divided between buyers and sellers

Elasticity in Areas Other Than Price

By the end of this section, you will be able to:

- Calculate the income elasticity of demand and the cross-price elasticity of demand
- Calculate the elasticity in labor and financial capital markets through an understanding of the elasticity of labor supply and the elasticity of savings
- Apply concepts of price elasticity to real-world situations

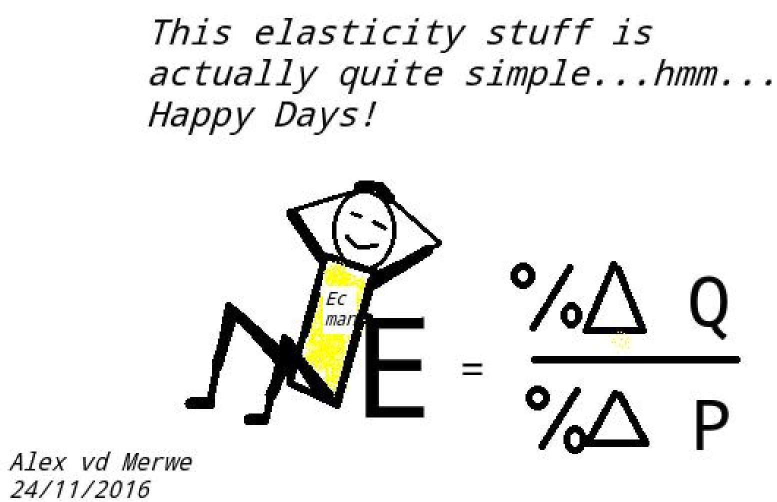


Figure 1.

The basic idea of elasticity—how a percentage change in one variable causes a percentage change in another variable—does not apply only to the responsiveness of supply and demand to changes in the price of a product. Recall that quantity demanded (Qd) depends on income, tastes and preferences, the prices of related goods, and so on, as well as price. Similarly, quantity supplied (Qs) depends on the cost of production, and so on, as well as price. Elasticity can be measured for any determinant of supply and demand, not just the price.

Income Elasticity of Demand

The income elasticity of demand (Ey) is the percentage change in quantity demanded divided by the percentage change in income (Y). So this elasticity coefficient measures the responsiveness of quantity demanded to changes in income.
Equation:

$$\text{Income elasticity of demand (Ey)} = \frac{\%\text{ change in quantity demanded}}{\%\text{ change in income}}$$

For most products, most of the time, the income elasticity of demand is positive: that is, a rise in income will cause an increase in the quantity demanded. This pattern is common enough that these goods are referred to as normal goods. However, for a few goods, an increase in income means that one might purchase less of the good; for example, those with a higher income might buy fewer burgers, because they are buying more steak instead, or those with a higher income might buy less cheap wine and more imported beer. When the income elasticity of demand is negative, the good is called an inferior good.

The concepts of normal and inferior goods were introduced in the Demand and Supply model. A higher level of income for a normal good causes a demand curve to shift to the right for a normal good, which means that the income elasticity of demand is positive. How far the demand shifts depends on the income elasticity of demand. A higher income elasticity means a larger shift. However, for an inferior good, that is, when the income elasticity of demand is negative, a higher level of income would cause the demand curve for that good to shift to the left. Again, how much it shifts depends on how large the (negative) income elasticity is.

Income elasticity (Ey) is interpreted according to the same scale as that used to interpret price elasticity (Ep). If the absolute (i.e. ignoring negative values) value of Ey exceeds 1, demand is income elastic. If the absolute value of Ey is less than 1 demand is income inelastic and if the absolute value of Ey = 1, demand is income unit elastic.

Cross-Price Elasticity of Demand

A change in the price of one good can shift the quantity demanded for another good. If the two goods are complements, like bread and peanut butter, then a drop in the price of one good will lead to an increase in the quantity demanded of the other good. However, if the two goods are substitutes, like plane tickets and train tickets, then a drop in the price of one good will cause people to substitute toward that good, and to reduce consumption of the other good. Cheaper plane tickets lead to fewer train tickets, and vice versa.

The **cross-price elasticity of demand** (Ec) puts some meat on the bones of these ideas. The term “cross-price” refers to the idea that the price of one good is affecting the quantity demanded of a different good. Specifically, the cross-price elasticity of demand is the percentage change in the quantity of good A that is demanded as a result of a percentage change in the price of good B.
Equation:

$$\text{Cross-price elasticity of demand (Ec)} = \frac{\text{\% change in Qd of good A}}{\text{\% change in price of good B}}$$

Substitute goods have positive cross-price elasticities of demand: if good A is a substitute for good B, like coffee and tea, then a higher price for B will mean a greater quantity consumed of A. Complement goods have negative cross-price elasticities: if good A is a complement for good B, like coffee and sugar, then a higher price for B will mean a lower quantity consumed of A.

Cross-price elasticity (Ec) is interpreted according to the same scale as that used to interpret price elasticity (Ep) and income elasticity (Ey). If the absolute (i.e. ignoring negative values) value of Ec exceeds 1, demand for A is cross-price elastic with respect to change in price of related good B. If the absolute value of Ec is less than 1, then demand for A is cross-price inelastic with respect to change in price of related good B, and if the absolute value of Ey = 1, then demand for A is cross-price unit elastic with respect to change in price of related good B.

Expanding the Concept of Elasticity

The elasticity concept does not even need to relate to a typical supply or demand curve at all. For example, imagine that you are studying whether the Receiver of Revenue should spend more money on auditing tax returns. The question can be framed in terms of the elasticity of tax collections with respect to spending on tax enforcement; that is, what is the percentage change in tax collections derived from a percentage change in spending on tax enforcement?

With all of the elasticity concepts that have just been described, some of which are listed in Table 1, the possibility of confusion arises. When you hear the phrases “elasticity of demand” or “elasticity of supply,” they refer to the elasticity with respect to price. Sometimes, either to be extremely clear or because a wide variety of elasticities are being discussed, the elasticity of demand or the demand elasticity will be called the price elasticity of demand or the “elasticity of demand with respect to price.” Similarly, elasticity of supply or the supply elasticity is sometimes called, to avoid any possibility of confusion, the price elasticity of supply or “the elasticity of supply with respect to price.” But in whatever context elasticity is spoken about, the idea always refers to percentage change in one variable, almost always a price or money variable, and how it causes a percentage change in another variable, typically a quantity variable of some kind.

$\text{Income elasticity of demand} = \frac{\%\text{ change in Qd}}{\%\text{ change in income}}$
$\text{Cross-price elasticity of demand} = \frac{\%\text{ change in Qd of good A}}{\%\text{ change in price of good B}}$
$\text{Wage elasticity of labor supply} = \frac{\%\text{ change in quantity of labor supplied}}{\%\text{ change in wage}}$
$\text{Wage elasticity of labor demand} = \frac{\%\text{ change in quantity of labor demanded}}{\%\text{ change in wage}}$
$\text{Interest rate elasticity of savings} = \frac{\%\text{ change in quantity of savings}}{\%\text{ change in interest rate}}$
$\text{Interest rate elasticity of borrowing} = \frac{\%\text{ change in quantity of borrowing}}{\%\text{ change in interest rate}}$

Formulas for Calculating Elasticity

Key Concepts and Summary

Elasticity is a general term, referring to percentage change of one variable divided by percentage change of a related variable that can be applied to many economic connections. For instance, the income elasticity of demand is the percentage change in quantity demanded divided by the percentage change in income. The cross-price elasticity of demand is the percentage change in the quantity demanded of a good divided by the percentage change in the price of another good. Elasticity applies in labor markets and financial capital markets just as it does in markets for goods and services. The wage elasticity of labor supply is the percentage change in the quantity of hours supplied divided by the percentage change

in the wage. The elasticity of savings with respect to interest rates is the percentage change in the quantity of savings divided by the percentage change in interest rates.

Self-Check Questions

Exercise:

Problem: What would the petrol price elasticity of supply mean to taxi bosses?

Solution:

The percentage change in quantity supplied as a result of a given percentage change in the price of petrol.

Exercise:

Problem:

The average annual income rises from R375,000 to R570,000, and the quantity of bread consumed in a year by the average person falls from 30 loaves to 22 loaves. What is the income elasticity (using the point elasticity formula) of bread consumption? Is bread a normal or an inferior good?

Solution:
Equation:

$$\begin{aligned}\text{Percentage change in quantity demanded} &= [(\text{change in quantity})/(\text{original quantity})] \times 100\\ &= [22 - 30]/[30] \times 100\\ &= -8/30 \times 100\\ &= -26.67\\ \text{Percentage change in income} &= [(\text{change in income})/(\text{original income})] \times 100\\ &= [570{,}000 - 375{,}000]/[375{,}000] \times 100\\ &= 195{,}000/375{,}000 \times 100\\ &= 52.00\end{aligned}$$

Ey = % change in Qd/% change in income = -26.67%52% = -0.51. In this example, bread is an inferior good because its consumption falls as income rises. Also the absolute value of Ey (0.51) is less than 1 so demand for bread is also income inelastic (not very sensitive to income changes).

Exercise:

Problem:

Suppose the cross-price elasticity of apples with respect to the price of oranges is 0.4, and the price of oranges falls by 3%. What will happen to the demand for apples assuming apples and oranges are regarded as substitutes?

Solution:

The formula for cross-price elasticity is % change in Qd for apples / % change in P of oranges. Multiplying both sides by % change in P of oranges yields:

% change in Qd for apples = cross-price elasticity X% change in P of oranges

= 0.4 × (3%) = 1.2%, or a 1.2 % increase in quantity demanded of apples (positive Ec indicates a substitute relationship between apples and oranges).

Review Questions

Exercise:

Problem: What is the formula for the income elasticity of demand?

Exercise:

Problem: What is the formula for the cross-price elasticity of demand?

Critical Thinking Questions

Exercise:

Problem:

Normal goods are defined as having a positive income elasticity. We can divide normal goods into two types: those whose income elasticity is less than one and those whose income elasticity is greater than one. Think about products that would fall into each category. Can you come up with a name for each category?

Exercise:

Problem:

Suppose you could buy shoes one at a time, rather than in pairs. What do you predict the cross-price elasticity for left shoes and right shoes would be?

Glossary

cross-price elasticity of demand
: the percentage change in the quantity of good A that is demanded as a result of a percentage change in good B

elasticity of savings
: the percentage change in the quantity of savings divided by the percentage change in interest rates

wage elasticity of labor supply
: the percentage change in hours worked divided by the percentage change in wages

Introduction to Cost and Industry Structure
class="introduction"

Amazon is an American international electronic commerce company that sells books, among many other things, shipping them directly to the consumer. There is no brick-and-mortar Amazon store. (Credit: modification of work by William Christiansen/Flickr Creative Commons)

Note:
Amazon
In less than two decades, Amazon.com has transformed the way books are sold, bought, and even read across the world. Prior to Amazon, books were primarily sold through independent bookstores with limited stocks such as Adams, Juta, Van Schaik and Exclusive Books. In the last decade, however, independent bookstores have become few and far between, some have gone out of business, and the remaining ones have had to reinvent their business models and cut costs to remain competitive. Online delivery (such as this free open economics textbook) and purchase of copyright books has indeed overtaken the more traditional business models built around the sale of bound print copies. How has Amazon changed the book selling industry? How has it managed to crush its competition?
A major reason for the giant retailer’s success is its production model and cost structure, which has enabled Amazon to undercut the prices of its competitors even when factoring in the cost of shipping. Read on to see how firms great (like Amazon) and small (like your corner cafe) determine what to sell, at what output and price.

Note:
Introduction to Cost and Industry Structure
In this chapter, you will learn about:

- Explicit and Implicit Costs, and Accounting and Economic Profit
- The Structure of Costs in the Short Run
- The Structure of Costs in the Long Run

This chapter is the first of four chapters that explore the *theory of the firm*. This theory explains that firms behave in much the same way as consumers behave. What does that mean? Let’s define what is meant by the firm. A **firm** (or business) combines inputs of labor, capital, land, and raw or finished component materials to produce outputs. If the firm is successful, the outputs are more valuable than the inputs. This activity of **production** goes beyond manufacturing (i.e., making things). It includes any process or service that creates value, including transportation, distribution, wholesale and retail sales. Production involves a number of important decisions that define the behavior of firms. These decisions include, but are not limited to:

- What product or products should the firm produce?
- How should the products be produced (i.e., what production process should be used)?
- How much output should the firm produce?
- What price should the firm charge for its products?
- How much labor should the firm employ?

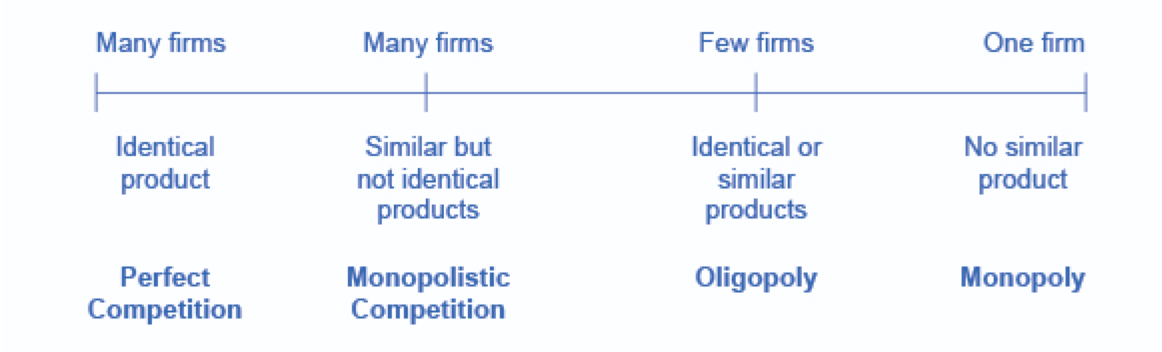
The answers to these questions depend on the production and cost conditions facing each firm. The answers also depend on the structure of the market for the product(s) in question. Market structure is a multidimensional concept that involves how competitive the industry is. It is defined by questions such as these:

- How much market power does each firm in the industry possess?
- How similar is each firm’s product to the products of other firms in the industry?

- How difficult is it for new firms to enter the industry?
- Do firms compete on the basis of price, advertising, or other product differences?

Figure 2 illustrates the range of different market structures, which we will explore in Perfect Competition, Monopoly, and Monopolistic Competition and Oligopoly.

The Spectrum of Competition



Firms face different competitive situations. At one extreme—perfect competition—many firms are all trying to sell identical products. At the other extreme—monopoly—only one firm is selling the product, and this firm faces no competition. Monopolistic competition and oligopoly fall between the extremes of perfect competition and monopoly. Monopolistic competition is a situation with many firms selling similar, but not identical, products. Oligopoly is a situation with few firms that sell identical or similar products.

First let's take a look at how firms determine their costs and desired profit levels. Then we will discuss costs in the short run and long run and the factors that can influence each.

Explicit and Implicit Costs, and Accounting and Economic Profit

By the end of this section, you will be able to:

- Explain the difference between explicit costs and implicit costs
- Understand the relationship between cost and revenue

Private enterprise, the ownership of businesses by private individuals, is a characteristic of the South African economy. When people think of businesses, often giants like Standard Bank, Sasol, MTN and Toyota SA come to mind. But firms come in all sizes. In fact the majority of private sector firms in South Africa employ less than 10 people (Schussler: 2012) which means that collectively they employ more people than the bigger firms. The biggest single cost for most firms is that of labor.

Table 1: Employment and earnings by industry		
Industry	**December 2015 (thousands)**	**Earnings (Rmillion)**
Mining	462	29 053
Manufacturing	1 129	61 726
Electricity	58	7 447
Construction	471	21 974
Trade	1 896	75 530
Transport	435	31 664
Business Services	2 009	130 665
Community Services	2 532	168 667
Total	8 992	526 726

Source: Statistics South Africa, 2016

Each of the businesses in these sectors, regardless of size or complexity, tries to earn a profit:

Equation:

$$\text{Profit} = \text{Total Revenue} - \text{Total Cost}$$

Total **revenue** is the income brought into the firm from selling its products. It is calculated by multiplying the price of the product times the quantity of output sold:

Equation:

$$\text{Total Revenue} = \text{Price} \times \text{Quantity}$$

We will see in the following chapters that revenue depends on the demand for the firm's products.

We can distinguish between two types of cost: explicit and implicit. **Explicit costs** are out-of-pocket costs, that is, money payments that are actually made. Wages that a firm pays its employees or rent that a firm pays for its office are explicit costs. Explicit costs can also be termed accounting costs. **Implicit costs** are not as obvious or clear as explicit costs, but just as important. Implicit costs represent the opportunity cost of using resources already owned by the firm. Often for small businesses, they are resources contributed by the owners; for example, working in the business while not getting a formal salary, or using the garage or spare room at home as a retail store. Implicit costs also allow for depreciation of goods, materials, and equipment that are necessary for a company to operate. (See the Work it Out feature for an extended example.)

These two definitions of cost are important for distinguishing between two conceptions of profit, accounting profit and economic profit. **Accounting profit** is a cash concept. It means total revenue minus explicit costs—the difference between Rands brought in and Rands paid out. **Economic profit** is total revenue minus total cost, including both explicit and implicit costs. The difference is important because even though a business pays income taxes (or company tax) based on its accounting profit, whether or not it is economically successful depends on its economic profit.

Note:
Calculating Implicit Costs

Consider the following example. Fred currently works for a big law firm. He is considering opening his own legal practice, where he expects to earn R750,000 per year once he gets established. To run his own firm, he would need an office and a law clerk. He has found the perfect office in Sandton, which rents for R60,000 per year. A law clerk could be hired for R120,000 per year. If these figures are accurate, would Fred's legal practice be profitable?

Step 1. First you have to calculate the costs. You can take what you know about explicit costs and total them:

Equation:

$$\begin{array}{lr} \text{Office rental :} & \text{R}60{,}000 \\ \text{Law clerk's salary :} & +\text{R}120{,}000 \\ \hline \text{Total explicit costs :} & \text{R}180{,}000 \end{array}$$

Step 2. Subtracting the explicit costs from the revenue gives you the accounting profit.

Equation:

$$\begin{array}{lr} \text{Revenues :} & \text{R}750{,}000 \\ \text{Explicit costs :} & -\text{R}180{,}000 \\ \hline \text{Accounting profit :} & \text{R}570{,}000 \end{array}$$

But these calculations consider only the explicit costs. To open his own practice, Fred would have to quit his current job, where he is earning an annual salary of R580,000. This would be an implicit cost (opportunity cost) of opening his own firm.

Step 3. You need to subtract both the explicit and implicit costs to determine the true economic profit:

Equation:

$$\begin{aligned} \text{Economic profit} &= \text{total revenues} - \text{explicit costs} - \text{implicit costs} \\ &= \text{R}750{,}000 - \text{R}180{,}000 - \text{R}580{,}000 \\ &= -\text{R}10{,}000 \text{ per year} \end{aligned}$$

Fred would be losing R10,000 per year. That does not mean he would not want to open his own business, but it does mean he would be earning R10,000 less than if he worked for the corporate firm.
Implicit costs can include other things as well. Maybe Fred values his leisure time, and starting his own firm would require him to put in more hours than at the corporate firm. In this case, the lost leisure would also be an implicit cost that would subtract from economic profits.

Now that we have an idea about the different types of costs, let's look at cost structures. A firm's cost structure in the long run (the future when nothing is fixed and everything can change) may be different from that in the short run (the present when some things cannot be changed and we just have to cope). We turn to that distinction in the next section.

Key Concepts and Summary

Privately owned firms are motivated to earn profits. Profit is the difference between revenues and costs. While accounting profit considers only explicit costs, economic profit considers both explicit and implicit costs.

Self-Check Questions

Exercise:

Problem:

A firm had sales revenue of R1 million last year. It spent R600,000 on labor, R150,000 on capital and R200,000 on materials. What was the firm's accounting profit?

Solution:

Accounting profit = total revenues minus explicit costs =R1,000,000 – (R600,000 + R150,000 + R200,000) = R50,000.

Exercise:

Problem:

Continuing from Exercise 1, the firm's factory sits on land owned by the firm that could be rented out for R30,000 per year. What was the firm's economic profit last year?

Solution:

Economic profit = accounting profit minus implicit cost = R50,000 – R30,000 = R20,000.

Review Questions

Exercise:

Problem: What are explicit and implicit costs?

Exercise:

Problem:

Would an interest payment on a loan to a firm be considered an explicit or implicit cost?

Exercise:

Problem:

What is the difference between accounting and economic profit?

Critical Thinking Questions

Exercise:

Problem:

Some small family owned businesses such as corner cafés or shops, sometimes exist even though they do not earn economic profits. How can you explain this?

Problems

Exercise:

Problem:

A firm is considering an investment that will earn a 6% rate of return. If it were to borrow the money, it would have to pay 8% interest on the loan, but it currently has the cash, so it will not need to borrow. Should the firm make the investment? Show your work.

References:
Schussler, M. 2012. Business owners in SA. Who, where, how big, what they do, and a few other facts. Available http://www.absa.co.za/deployedfiles/Absacoza/PDFs/Media%20Releases/Absa%20SME%20Index%20Launch.pdf (Accessed: 24 Apr 2016)

Statistics South Africa. 2016. Quarterly employment statistics (QES) December 2015. Statistical release, no. PO277. Pretoria: Statistics South Africa. Available http://www.statssa.gov.za/ (Accessed: 24 Apr 2016)

Glossary

accounting profit
: total revenues minus explicit costs, including depreciation

economic profit
: total revenues minus total costs (explicit plus implicit costs)

explicit costs

out-of-pocket costs for a firm, for example, payments for wages and salaries, rent, or materials

firm
: an organization that combines inputs of labor, capital, land, and raw or finished component materials to produce outputs.

implicit costs
: opportunity cost of resources already owned by the firm and used in business, for example, expanding a factory onto land already owned

private enterprise
: the ownership of businesses by private individuals

production
: the process of combining inputs to produce outputs, ideally of a value greater than the value of the inputs

revenue
: income from selling a firm’s product; defined as price times quantity sold

The Structure of Costs in the Short Run

By the end of this section, you will be able to:

- Analyze short-run costs as influenced by total cost, fixed cost, variable cost, marginal cost, and average cost.
- Calculate average profit
- Evaluate patterns of costs to determine potential profit

The cost of producing a firm's output depends on how much labor, physical capital and **materials** the firm uses. A list of the costs involved in producing cars will look very different from the costs involved in producing computer software or haircuts or fast-food meals. However, the cost structure of all firms can be broken down into some common underlying patterns. When a firm looks at its **total costs** of production in the short run, a useful starting point is to divide total costs into two categories: fixed costs that cannot be changed in the short run and variable costs that can be changed.

The production function

Any production process (or function) requires inputs of land (natural resources), labour, physical capital and enterprise. Different production processes require different mixes of inputs. Inputs can generally be categorised as either **fixed** or **variable** . The amounts of fixed inputs do not change as production levels change. Fixed inputs give rise to **fixed costs** . The amount of variable inputs does change as production levels change. Variable inputs give rise to **variable costs** .

In order to manage a production process profitably, it is important to be able to understand the nature, pattern and behavior of the costs incurred by the business or firm. The process of production generates costs. So costs are dependent on the nature (type) of production which, in turn, determines the mix of inputs employed by the firm. Essentially, we need to understand the patterns and laws that govern production processes before we can better understand cost patterns and behaviors.

Vuyo's Gents Hair Salon

Let's take Vuyo's Gents hair salon in Edendale as an example of a typical business operation or process. Although it is relatively small and is a service business, it is generally representative of all types of businesses or firms. We will focus only on his main inputs of labor (variable input), rental (fixed input) and equipment (fixed input). Any production process, including Vuyo's, uses inputs to produce outputs. Vuyo needs to know how his costs will respond when he increases or decreases production (number of haircuts per day) since this will affect his profits. So he actually needs to measure how costs will be affected at different levels of production.

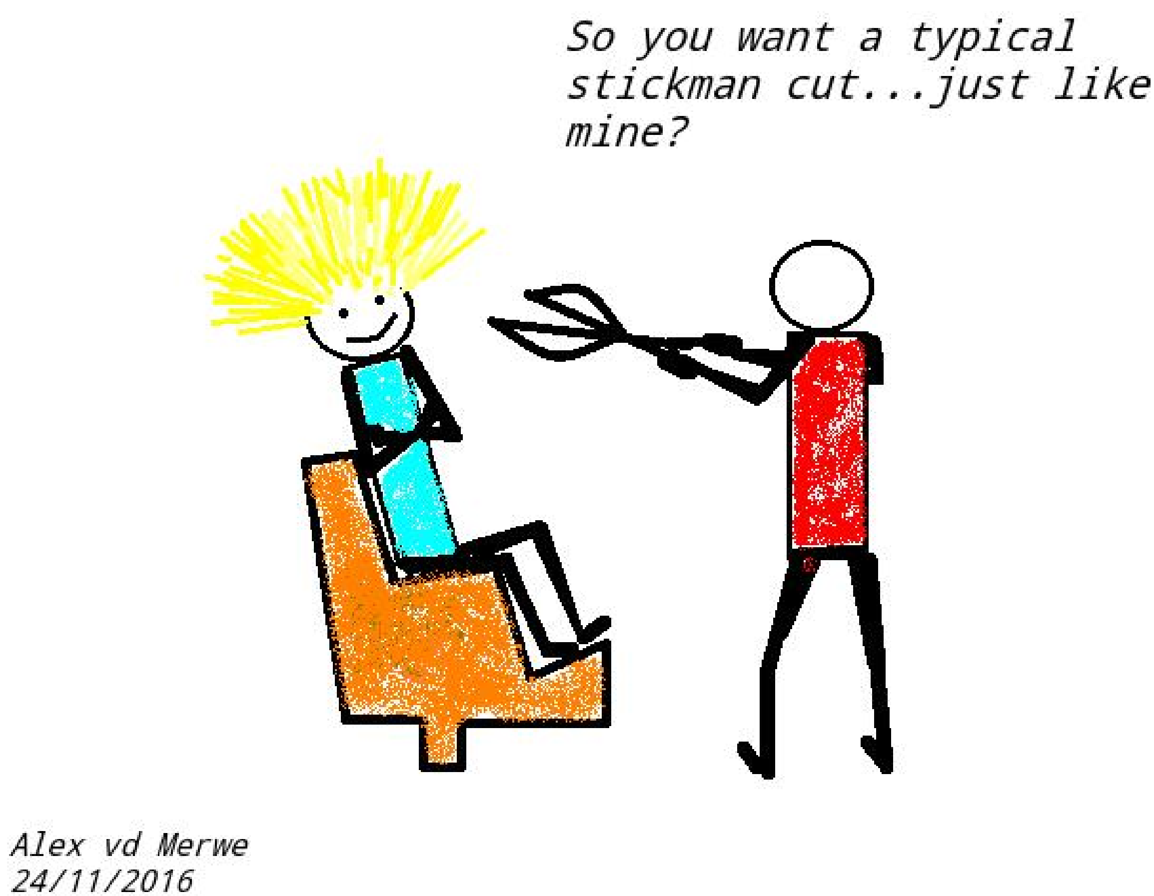


Figure 1

A good starting point in terms of measuring costs would be to start by measuring production. Producers like to look at issues and questions from different angles. One question might be: what is the total cost of this level of production? In such a case it would be necessary to measure **total production or TP** (total number of haircuts per day in Vuyos's case). A second question might be: what is the cost per unit (per haircut)? In this instance it would be appropriate to measure the production per worker (number of haircuts per hairdresser in Vuyo's example) the **average production or AP** . Yet another question might be: how will costs change if production levels change (either increase or decrease)? In this case we would need to measure **marginal production or MP** (the change in total production (number of haircuts) that results when there is a change in inputs (the employment of additional hairdressers).

So not only is it important to be able to measure production but it is equally important to be able to measure it in **different** ways in order to be able to answer different questions and to gain clearer insight into the behavior of costs and therefore business profits. There is no formula to measure total production. It is simply what it is: the total product, TP. Average product (AP) is total product divided by the variable input (usually labor input). So AP = TP ÷ quantity of variable input. Marginal product (MP) is the change in total product divided by the change in variable input (usually labor input). So MP = change in TP ÷ change in variable input.

The Short Run versus Long Run Production Periods
The theory of the firm distinguishes between the **"short run"** and the **"long run"** . These different "runs" refer to production time periods which are not specified in days, weeks, months or years. The short run (or the present) is simply the time period in which at least one input is fixed. The long run (or the future), on the other hand, is the time period in which all inputs are variable. The production strategies and decisions that firms take will differ depending on whether it is the "short run" or the "long run".

So let's take a look at the Vuyo's Gents Hair Salon production data presented in the table and graph below. Some interesting trends become apparent. Firstly, TP rises up to a point and then falls. This gives us our first clue about how costs will respond as production increases. Rising total

production (TP) should certainly cause **total cost** to increase. The production per worker (AP) also rises to a point and then falls. AP reflects how efficient individual workers are. So rising AP should result in initially falling **unit or average cost** and then as AP starts to fall (indicating declining worker efficiency), unit or average cost should start to rise.

Table 1: Vuyo's Gents Hair Salon production data

Labour (Hairdressers)	Total production, TP (Haircuts/day)	Marginal production, MP (Haircuts/day)	Average production, AP (Haircuts/day)
0	**0**		**/**
		16	
1	16		16
		24	
2	40		20
		20	
3	60		20
		12	
4	72		18
		8	
5	80		16
		4	
6	84		14
		-2	
7	82		11.7

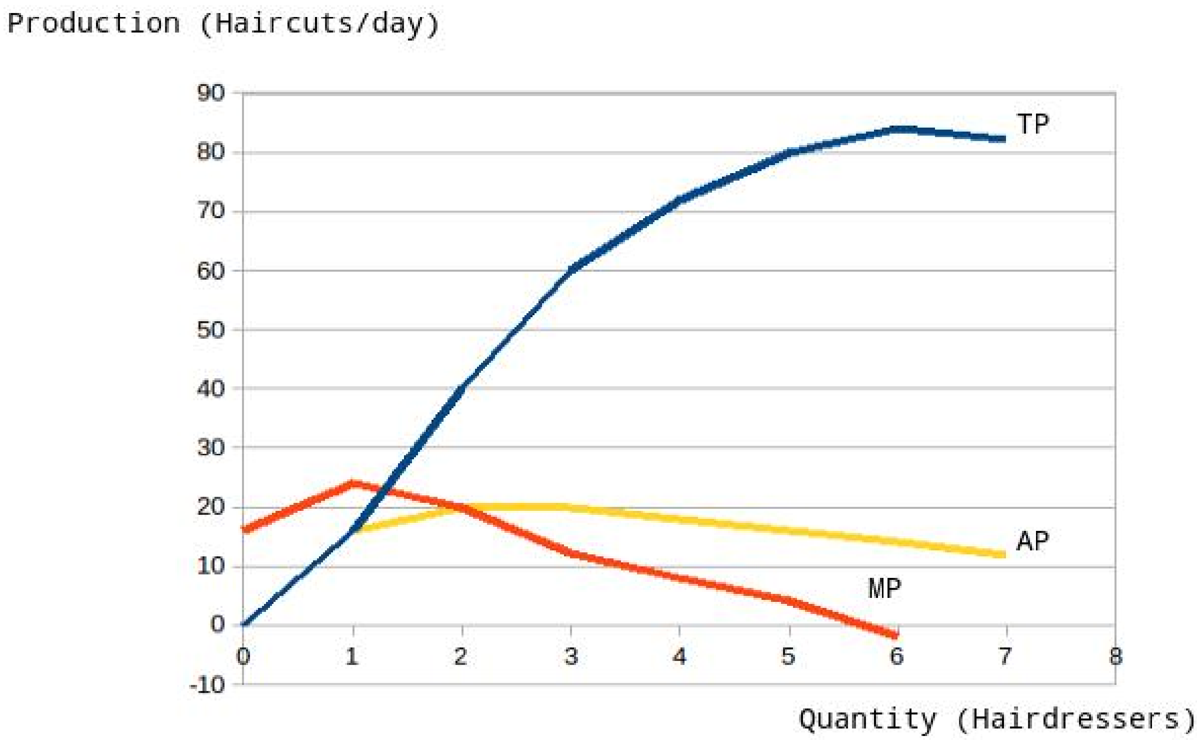


Figure 2. Total, Average and Marginal Production.

MP follows a similar pattern to AP, rising to a point and then declining. What does this indicate? Up until some point each new worker employed in the production process will increase total production by increasing amounts. Again, this reflects greater efficiency of workers as more are hired and the teamwork in the business improves. One would expect the additional (marginal) cost of hiring additional workers to fall as MP rises because teamwork improves efficiency. Of course, the team can get too big and employing more workers can lead to overcrowding and loss of focus of employees. So inefficiency sets in and MP starts to decline and so the additional (marginal) cost of employing additional workers will start to rise.

The total product, average product and marginal product curves are "mirrors" of the total cost, average cost and marginal cost curves (which we will see later) in that they reflect the shapes of these curves (although in reverse). The shapes of the TP, AP and MP curves as they each rise to a

maximum and then fall suggest that there is an ideal combination or mix of fixed and variable inputs that a business or firm should employ. So, for example, if Vuyo employs too few hairdressers (variable input) for his shop and equipment (fixed inputs), his productivity and profitability will not be good. Similarly, if he employs too many hairdressers some may end up lazing around or chatting so that productivity will fall. These patterns that all production processes follow are described by the law of diminishing returns.

Production in the short run
The law of diminishing (marginal) returns predicts that, as more of a variable input is combined with a fixed input, there will come a point when each extra unit of the variable input will produce less extra output than the previous unit. This, then, is the law that governs production in the present, that is, in the short run (when there is at least one fixed input). So this law of production forces businesses to find the best or ideal combinations of variable and fixed inputs to minimise costs and so maximise profitability. In the short run, firms can change their production levels only by changing their variable inputs. They cannot change their fixed inputs.

Fixed and Variable Costs

Fixed costs, being the result of fixed inputs, are expenditures that do not change regardless of the level of production, at least not in the short term. Whether you produce a lot or a little, the fixed costs are the same. One example is the rent on Vuyo's salon or a factory or a retail space. Once you sign the lease, the rent is the same regardless of how much you produce, at least until the lease runs out. Fixed costs can take many other forms: for example, the cost of machinery or equipment to produce the product, research and development costs to develop new products, even an expense like advertising to popularize a brand name. The level of fixed costs varies according to the specific line of business: for instance, manufacturing computer chips requires an expensive factory, but a local transport business, for example, can get by with almost no fixed costs at all if it rents trucks by the day when needed.

Variable costs, on the other hand, are incurred in the act of producing—the more you produce, the greater the variable cost since more variable inputs are required to increase production. Labor is treated as a variable cost, since producing a greater quantity of a good or service typically requires more workers or more work hours. Variable costs would also include raw materials.

As a concrete example of fixed and variable costs, let's go back to Vuyo's Gents Hair Salon. The data for output and costs are shown in the table below. The fixed costs of operating the Salon, including the space and equipment, are R160 per day. The variable costs are the costs of hiring hairdressers, which in our example is R80 per hairdresser each day. The first two columns of the table show the quantity of haircuts the salon can produce as it hires additional hairdressers. The third column shows the fixed costs, which do not change regardless of the level of production. The fourth column shows the variable costs at each level of output. These are calculated by taking the amount of labor hired and multiplying by the wage. For example, two hairdressers cost: 2 × R80 = R160. Adding together the fixed costs in the third column and the variable costs in the fourth column produces the total costs in the fifth column. So, for example, with two hairdressers the total cost is: R160 + R160 = R320.

Table 2: Vuyo's Gents Hair Salon cost data

Labor (Hairdressers)	Quantity	Fixed Cost (Rands)	Variable Cost (Rands)	Total Cost (Rands)
1	16	160	80	240
2	40	160	160	320
3	60	160	240	400
4	72	160	320	480
5	80	160	400	560
6	84	160	480	640
7	82	160	560	720

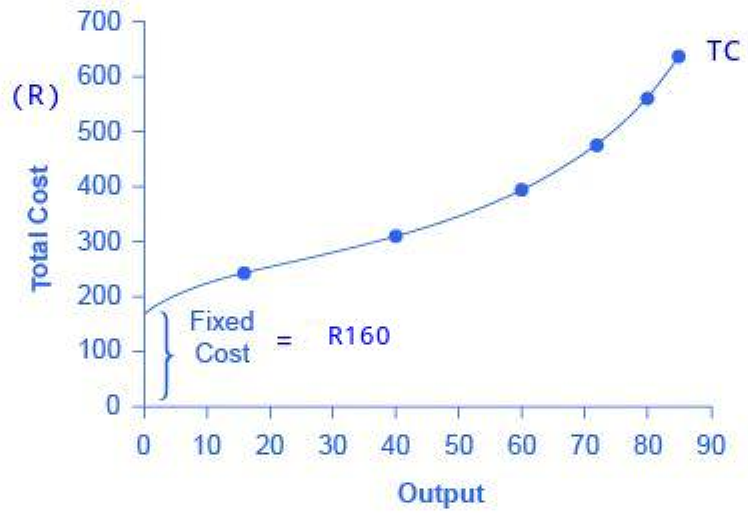


Figure 3. At zero production, the fixed costs of R160 are still present. As production increases, variable costs are added to fixed costs, and the total cost is the sum of the two.

The relationship between the quantity of output being produced and the cost of producing that output is shown graphically in the figure. The fixed costs are always shown as the vertical intercept of the total cost curve; that is, they are the costs incurred when output is zero so there are no variable costs.

You can see from the graph that once production starts, total costs and variable costs rise. We kind of expected this from examining the total product (TP) curve earlier. It is logical that total costs will rise as production rises. In the short run the cause of rising total cost is total variable cost since total fixed cost, by definition, does not change. While variable costs may initially increase at a decreasing rate, at some point they begin increasing at an increasing rate. This is caused by diminishing

marginal returns, discussed earlier in the chapter and also in relation to the PPC (Production Possibilities Curve) in a previous chapter. Thus, as the number of hairdressers increases from zero to one in the table, output increases from 0 to 16 for a marginal gain of 16; as the number rises from one to two hairdressers, output increases from 16 to 40, a marginal gain of 24. From that point on, though, the marginal gain in output diminishes as each additional hairdresser is added. For example, as the number of hairdressers rises from two to three, the marginal output gain is only 20; and as the number rises from three to four, the marginal gain is only 12.

To understand the reason behind this pattern, consider that a one-man salon is a very busy operation. The single hairdresser needs to do everything: say hello to people entering, answer the phone, cut hair, sweep up, and run the cash register. A second hairdresser reduces the level of disruption from jumping back and forth between these tasks, and allows a greater division of labor and specialization. The result can be greater increasing marginal returns (more efficiency). However, as other hairdressers are added, the advantage of each additional hairdresser is less, since the specialization of labor can only go so far. The addition of a sixth or seventh or eighth hairdresser just to greet people at the door will have less impact than the second one did (the business becomes less efficient). This is the pattern of diminishing marginal returns which we learned about earlier in the chapter when we were looking at short run production patterns. As a result, the total costs of production will begin to rise more rapidly as output increases. At some point, you may even see negative returns as the additional hairdressers begin bumping elbows, knocking into customers and generally getting in each others way. In this case, the addition of still more hairdressers would actually cause output to decrease.

This pattern of diminishing marginal returns is common in production. As another example, consider the problem of irrigating a crop on a farmer’s field. The plot of land is the fixed factor of production, while the water that can be added to the land is the key variable cost. As the farmer adds water to the land, output increases. But adding more and more water brings smaller and smaller increases in output, until at some point the water floods the field and actually reduces output. Recall that diminishing marginal returns occur because, at a given level of fixed costs (and therefore fixed

inputs), each additional variable input contributes less and less to overall production. Remember, also, that in the short run businesses can change their output or production levels only by changing the amount of variable input. They cannot change the amount of fixed input immediately (the present which is also known as the short run). Which means that all production in the short run is governed by the law of diminishing returns. This law applies only in the short run (i.e. when there is at least one fixed input).

Average Total Cost, Average Variable Cost, Marginal Cost

The breakdown of total costs into fixed and variable costs can provide a basis for other insights as well. The first five columns of Table 3 below duplicate Table 2, but the last three columns show average total costs, average variable costs, and marginal costs. These new measures analyze costs on a per-unit (rather than a total) basis and are reflected in the curves shown in Figure 4.

Table 3: Alternative measures of cost

Labor (Hairdressers)	Quantity (Haircuts)	Fixed Cost (Rands)	Variable Cost (Rands)	Total Cost (Rands)	Marginal Cost (Rands)	Average Total Cost (Rands)	Average Variable Cost (Rands)
0	0	160	0	160		/	/
					5		
1	16	160	80	240		15	5
					3.3		
2	40	160	160	320		8	4
					4		
3	60	160	240	400		6.6	4
					6.6		
4	72	160	320	480		6.6	4.4
					10		
5	80	160	400	560		7	5
					20		
6	84	160	480	640		7.6	5.7

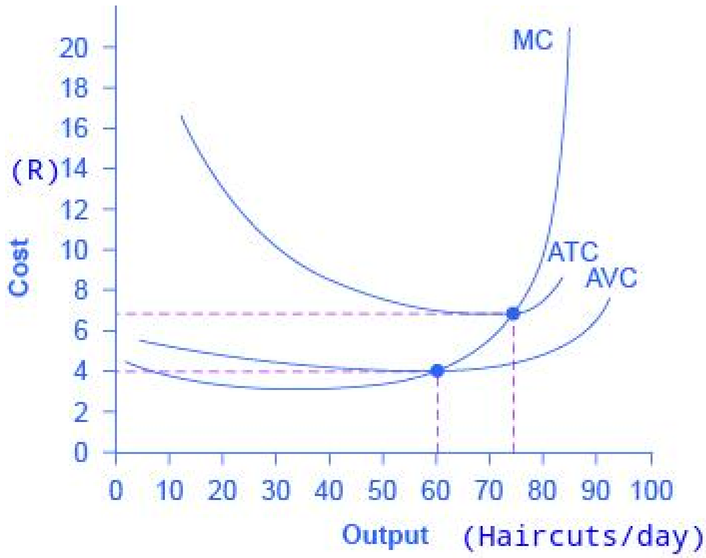


Figure 4. Average Total Cost, Average Variable Cost and Marginal Cost.

Average total cost (sometimes referred to simply as average cost) is total cost divided by the quantity of output. Since the total cost of producing 40 haircuts is R320, the average total cost for producing each of 40 haircuts is R320 divided by 40, or R8 per haircut. Average cost curves are typically U-shaped, as shown in the diagram above. Average total cost starts off relatively high, because at low levels of output total costs are dominated by the fixed cost; mathematically, the denominator is so small that average total cost is large. Average total cost then declines, as the fixed costs are spread over an increasing quantity of output. In the average cost calculation, the rise in the numerator of total costs is relatively small compared to the rise in the denominator of quantity produced. But as output expands still further, the average cost begins to rise. At the right side of the average cost curve, total costs begin rising more rapidly as diminishing returns kick in. Please note that the **cost** of a haircut is not the **price** of the haircut. We will look at pricing in the following chapters.

Average variable cost is obtained when variable cost is divided by quantity of output. For example, the variable cost of producing 80 haircuts is R400, so the average variable cost is R400 divided by 80, or R5 per haircut. Note

that at any level of output, the average variable cost curve will always lie below the curve for average total cost. The reason is that average total cost includes average variable cost and average fixed cost. Thus, for Q = 80 haircuts, the average total cost is R8 per haircut, while the average variable cost is R5 per haircut. However, as output grows, fixed costs become relatively less important (since they do not rise with output), so average variable cost sneaks closer to average cost. Again, please note that the **cost** of a haircut is not the **price** of the haircut. We will look at pricing in the following chapters.

Average total and variable costs measure the average costs of producing some quantity of output. Marginal cost is somewhat different. **Marginal cost** is the additional cost of producing one more unit of output. So it is not the cost per unit of *all* units being produced, but only the next one (or next few). Marginal cost can be calculated by taking the change in total cost and dividing it by the change in quantity of output. For example, as quantity produced increases from 40 to 60 haircuts, total costs rise by R400 – R320, or R80. Thus, the marginal cost for each of those marginal (additional) 20 units will be R80 divided by 20, or R4 per haircut. The marginal cost curve is generally upward-sloping, because diminishing marginal returns implies that additional units are more costly to produce. A small range of increasing marginal returns can be seen in the figure as a dip in the marginal cost curve before it starts rising. There is a point at which marginal and average costs meet, as the following Clear it Up feature discusses.

Note:

Where do marginal and average costs meet?

The marginal cost curve intersects the average cost line exactly at the bottom of the average cost curve—which occurs at a quantity of between 60 - 72 haircuts per day and at a total **per unit** cost (or average total cost) of R6.60 in the diagram above. The reason why the intersection occurs at this point is built into the economic meaning of marginal and average costs. If the marginal cost of production is below the average cost for producing previous units, as it is for the points to the left of where MC crosses ATC, then producing one more additional unit will reduce average

costs overall—and the ATC curve will be downward-sloping in this zone. Conversely, if the marginal cost of production for producing an additional unit is above the average cost for producing the earlier units, as it is for points to the right of where MC crosses ATC, then producing a marginal unit will increase average costs overall—and the ATC curve must be upward-sloping in this zone. The point of transition, between where MC is pulling ATC down and where it is pulling it up, must occur at the minimum point of the ATC curve.

This idea of the marginal cost "pulling down" the average cost or "pulling up" the average cost may sound complicated, but think about it in terms of your own test marks. If the score on the most recent online economics quiz you take is lower than your average score on previous quizzes, then the marginal (additional) quiz pulls down your average mark. If your score on the most recent quiz is higher than the average on previous quizzes, the marginal quiz pulls up your average. In this same way, low marginal costs of production first pull down average costs and then higher marginal costs pull them up.

The numerical calculations behind average cost, average variable cost, and marginal cost will change from firm to firm. However, the general patterns of these curves, and the relationships and economic intuition behind them, will not change.

Lessons from Alternative Measures of Costs

Breaking down total costs into fixed cost, marginal cost, average total cost, and average variable cost is useful because each statistic offers its own insights for the firm.

Whatever the firm's quantity of production, total revenue must exceed total costs if it is to earn a profit. Fixed costs are often sunk costs that cannot be recouped. In thinking about what to do next, sunk costs should typically be ignored, since this spending has already been made and cannot be changed. However, variable costs can be changed, so they convey information about

the firm's ability to cut costs in the present and the extent to which costs will increase if production rises.

Note:

Why are total cost and average cost not on the same graph?

Total cost, fixed cost, and variable cost each reflect different aspects of the cost of production over the entire quantity of output being produced. These costs are measured in Rands. In contrast, average cost (or average total cost, ATC), and average variable cost (AVC) are costs **per unit** . In the example of Vuyo's Gents Hair Salon they are measured as cost per haircut. Marginal cost (MC) is the **additional or extra cost** that results from additional or extra output. Thus, it would not make sense to put all of these numbers on the same graph, since they are measured in different units (total Rand cost versus Rand cost per unit of output, or Rand cost per unit versus additional cost per unit increase in production). Just like production (measured either as TP, MP or AP), costs can be measured differently (e.g.TC, MC or ATC). Why not use just one measure? The reason is that the different measures of production and cost are used to answer different questions. In the next chapters we will see how these different measures become useful.

It would be as if the vertical axis measured two different things. In addition, as a practical matter, if they were on the same graph, the lines for marginal cost, average cost, and average variable cost would appear almost flat against the horizontal axis, compared to the values for total cost, fixed cost, and variable cost. Using the figures from the previous example, the total cost of producing 40 haircuts is R320. But the average cost is R320 divided by 40, or R8 per haircut. If you graphed both total and average cost on the same axes, the average cost would hardly show. This is due to the large difference in the scales required to graph the total cost curves (large scale required) and the average and marginal cost curves (much smaller scale required).

Average cost tells a firm whether it can earn profits given the current price in the market. If we divide profit by the quantity of output produced we get

average profit, also known as the firm’s *profit margin*. Expanding the equation for profit gives:

Equation:

$$\begin{aligned}\text{average profit} &= \frac{\text{profit}}{\text{quantity produced}}\\ &= \frac{\text{total revenue} - \text{total cost}}{\text{quantity produced}}\\ &= \frac{\text{total revenue}}{\text{quantity produced}} - \frac{\text{total cost}}{\text{quantity produced}}\\ &= \text{average revenue} - \text{average cost}\end{aligned}$$

But note that:

Equation:

$$\begin{aligned}\text{average revenue} &= \frac{\text{price} \times \text{quantity produced}}{\text{quantity produced}}\\ &= \text{price}\end{aligned}$$

Thus:

Equation:

$$\text{average profit} = \text{price} - \text{average cost}$$

This is the firm’s profit margin. This definition implies that if the market price is above average cost, average profit, and thus total profit, will be positive; if price is below average cost, then profits will be negative.

The marginal cost of producing an additional unit can be compared with the marginal revenue gained by selling that additional unit to reveal whether the additional unit is adding to total profit—or not. Thus, marginal cost helps producers understand how profits would be affected by increasing or decreasing production.

A Variety of Cost Patterns

The pattern of costs varies among industries and even among firms in the same industry. Some businesses have high fixed costs, but low marginal costs. Consider, for example, an Internet company that provides medical advice to customers. Such a company might be paid by consumers directly, or perhaps hospitals or healthcare practices might subscribe on behalf of their patients. Setting up the website, collecting the information, writing the content, and buying or leasing the computer space to handle the web traffic are all fixed costs that must be undertaken before the site can work. However, once the website is up and running, it can provide a high quantity of service with relatively low variable costs, like the cost of monitoring the system and updating the information. In this case, the total cost curve might start at a high level, because of the high fixed costs, but then might appear close to flat, up to a large quantity of output, reflecting the low variable costs of operation. If the website is popular, however, a large rise in the number of visitors will overwhelm the website, and increasing output further could require a purchase of additional computer space.

For other firms, fixed costs may be relatively low. For example, consider a small garden service operating in the suburbs of Pietermaritzburg. This business could probably be run from the owner's home (so no rent to pay) and it would need need little more than a bakkie to transport workers, lawnmowers and rakes to the homes of customers. Still other firms may find that diminishing marginal returns set in quite sharply. If a manufacturing plant in the industrial area of Mkondeni in Pietermaritzburg tried to run 24 hours a day, seven days a week, little time remains for routine maintenance of the equipment, and marginal costs can increase dramatically as the firm struggles to repair and replace overworked equipment.

Every firm can gain insight into its task of earning profits by dividing its total costs into fixed and variable costs, and then using these calculations as a basis for average total cost, average variable cost, and marginal cost. However, making a final decision about the profit-maximizing quantity to produce and the price to charge will require combining these perspectives on cost with an analysis of sales and revenue, which in turn requires looking at the market structure in which the firm finds itself. Before we turn

to the analysis of market structure in other chapters, we will analyze the firm’s cost structure from a long-run perspective.

Key Concepts and Summary

In a short-run perspective, a firm’s total costs can be divided into fixed costs, which a firm must incur before producing any output, and variable costs, which the firm incurs in the act of producing. Fixed costs are sunk costs; that is, because they are in the past and cannot be altered, they should play no role in economic decisions about future production or pricing. Variable costs typically show diminishing marginal returns, so that the marginal cost of producing higher levels of output rises.

Marginal cost is calculated by taking the change in total cost (or the change in variable cost, which will be the same thing since fixed costs do not change) and dividing it by the change in output, for each possible change in output. Marginal costs are typically rising but can also be u-shaped. The u-shape of the MC curve is due to the law of diminishing returns. On the downward sloping section of the MC curve, the firm experiences increasing returns to the variable factor input as the firm improves the mix of variable with fixed inputs (efficiency is improving) . On the upward sloping section of the MC curve diminishing returns to the variable factor input are experienced as the firm begins to exceed the ideal amount of variable input that should be employed together with the fixed input (inefficiency starts to creep in). A firm can compare marginal cost to the additional revenue it gains from selling another unit to find out whether its marginal unit is adding to profit..

Average total cost is calculated by taking total cost and dividing by total output at each different level of output. Average costs are typically U-shaped on a graph. The u-shape of the ATC curve is also due to the law of diminishing returns. On the downward sloping section of the ATC curve, the firm experiences increasing returns to the variable factor input as the firm improves the mix of variable with fixed inputs (efficiency is improving) . On the upward sloping section of the ATC curve diminishing returns to the variable factor input are experienced as the firm begins to exceed the ideal amount of variable input that should be employed together

with the fixed input (inefficiency starts to creep in). If a firm's average cost of production is lower than the market price, a firm will be earning profits.

Average variable cost is calculated by taking variable cost and dividing by the total output at each level of output. Average variable costs are typically U-shaped. As in the cases of MC and ATC, the u-shape of the AVC curve is due to the law of diminishing returns. On the downward sloping section of the AVC curve, the firm experiences increasing returns to the variable factor input as the firm improves the mix of variable with fixed inputs (efficiency is improving) . On the upward sloping section of the AVC curve diminishing returns to the variable factor input are experienced as the firm begins to exceed the ideal amount of variable input that should be employed together with the fixed input (inefficiency starts to creep in). If a firm's average variable cost of production is lower than the market price, then the firm would be earning profits if fixed costs are left out of the picture.

Self-Check Questions

Exercise 1

Lebo is a full time student who runs two-hour short courses teaching basic word processing (Libreoffice and Word) to supplement her allowance. She rents computer laboratory space from the university for R300 per month which is her only fixed cost. Her variable costs are shown in the table below. Complete the table.

Table 4: Lebo's Costs

Q (Courses)	TVC (Rands)	TFC (Rands)	TC (Rands)	ATC (Rands)	AFC (Rands)	AVC (Rands)	MC (Rands)
0							
1	100						
2	250						
3	450						
4	700						
5	1000						
6	1350						

Table 5: Solution to Lebo's Cost Table

Q (Courses)	TVC (Rands)	TFC (Rands)	TC (Rands)	ATC (Rands)	AFC (Rands)	AVC (Rands)	MC (Rands)
0		300	300	/	/	/	
			0				100
1	100	300	400	400	300	100	
			0				150
2	250	300	550	275	150	125	
			0				200
3	450	300	750	250	100	150	
			0				250
4	700	300	1000	250	75	175	
			0				300
5	1000	300	1300	260	60	200	
			0				350
6	1350	300	1650	275	50	225	

Exercise 2
Based on your answers entered into Lebo's cost and revenue table, assume she runs 4 courses per month (which fits in with her lecture schedule) and she charges a flat rate of R255 per course (she does not charge per person). A maximum of 5 students may attend each course in which case each participant would pay R51 towards the total course price. If there were fewer participants each one would have to contribute more to make up the course price of R255.

- What will her profits or losses be?
- How can you tell at a glance whether she is making or losing money at this price by looking at average cost?
- At the given quantity (4 courses) and price (R255 per course), is the marginal unit (the 4th course in the month) produced adding to profits?

Solution

- Total revenue in this example will be a quantity of four courses multiplied by the price of R255 per course, which equals R1020. Total

cost when producing four courses is R1000. Thus, at this level of quantity and output Lebo experiences a small profit R20.

- If price is greater than average cost, the firm is making an economic (or pure) profit. At an output of four courses, the average cost is R250 per course whereas price is R255 per course. Thus, at a glance you can see Lebo is making a small profit.
- When producing four courses, marginal cost is R250 per course. Price is R255 per course. Thus, the marginal course (4th course in the month) is adding to profits, which suggests that Lebo could probably look at offering a 5th course, that is, increasing her production.

Review Questions

Exercise 3
What is the difference between fixed costs and variable costs?

Exercise 4
Are there fixed costs in the long-run? Explain briefly.

Exercise 5
Are fixed costs also sunk costs? Explain.

Exercise 6
What are diminishing marginal returns as they relate to costs?

Exercise 7
Which of the following costs are measured on a per-unit basis: fixed costs, average cost, average variable cost, variable costs, marginal cost?

Exercise 8
How is each of the following calculated: marginal cost, average total cost, average variable cost?

Critical Thinking Questions

Exercise 9

A common name for fixed cost is “overhead.” If you divide fixed cost by the quantity of output produced, you get average fixed cost. Suppose fixed cost is R1,000. What does the average fixed cost curve look like? Use your response to explain what “spreading the overhead” means.

Exercise 10
How does fixed cost affect marginal cost? Why is this relationship important?

Exercise 11
Average cost curves (except for average fixed cost) tend to be U-shaped, decreasing and then increasing. Marginal cost curves have the same shape, though this may be harder to see since most of the marginal cost curve is increasing. Why do you think that average and marginal cost curves have the same general shape?

Problems

Exercise 12
Return to Vuyo's Gents Hair Salon. What is the marginal gain in output from increasing the number of hairdressers from 4 to 5 and from 5 to 6? Does it continue the pattern of diminishing marginal returns?

Exercise 13
Compute the average total cost, average variable cost, and marginal cost of producing 60 and 72 haircuts. Draw the graph of the three curves between 60 and 72 haircuts.

Glossary

average profit
: profit divided by the quantity of output produced; profit margin

average total cost
: total cost divided by the quantity of output

average variable cost

variable cost divided by the quantity of output

fixed cost
: expenditure that must be made before production starts and that does not change regardless of the level of production

marginal cost
: the additional cost of producing one more unit

total cost
: the sum of fixed and variable costs of production

variable cost
: cost of production that increases with the quantity produced

The Structure of Costs in the Long Run

By the end of this section, you will be able to:

- Calculate total cost
- Identify economies of scale, diseconomies of scale, and constant returns to scale
- Interpret graphs of long-run average cost curves and short-run average cost curves
- Analyze cost and production in the long run and short run

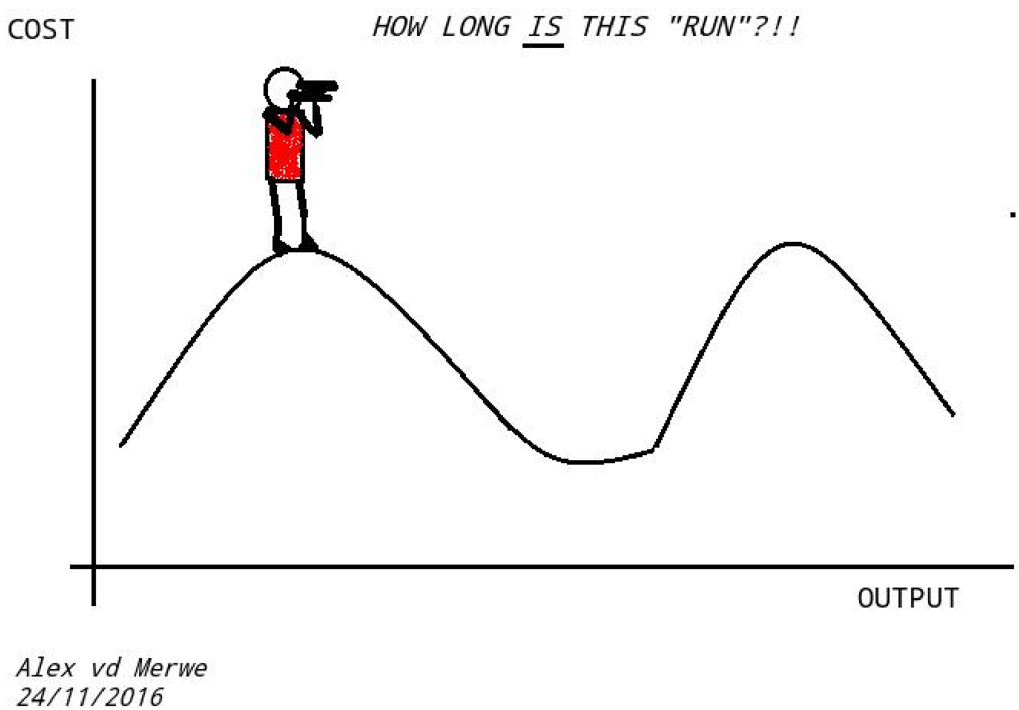


Figure 1. Econman has the wrong idea about what the "long run" means in economics.

The long run is the period of time when all costs are variable. The long run depends on the specifics of the firm in question—it is not a precise period

of time. If you have a one-year lease on your factory, then the long run is any period longer than a year, since after a year you are no longer bound by the lease. No costs are fixed in the long run. A firm can build new factories and purchase new machinery, or it can close existing facilities. In planning for the long run, the firm will compare alternative **production technologies** (or processes).

In this context, technology refers to all alternative methods of combining inputs to produce outputs. It does not refer to a specific new invention like the tablet computer. The firm will search for the production technology that allows it to produce the desired level of output at the lowest cost. After all, lower costs lead to higher profits—at least if total revenues remain unchanged. Moreover, each firm must fear that if it does not seek out the lowest-cost methods of production, then it may lose sales to competitor firms that find a way to produce and sell for less.

Choice of Production Technology

Many tasks can be performed with a range of combinations of labor and physical capital. For example, a firm can have human beings answering phones and taking messages, or it can invest in an automated voicemail system. A firm can hire file clerks and secretaries to manage a system of paper folders and file cabinets, or it can invest in a computerized recordkeeping system that will require fewer employees. A firm can hire workers to push supplies around a factory on rolling carts, it can invest in motorized vehicles, or it can invest in robots that carry materials without a driver. Firms often face a choice between buying many small machines, which need a worker to run each one, or buying one larger and more expensive machine, which requires only one or two workers to operate it. In short, physical capital and labor can often substitute for each other.

Consider the example of a private firm that is hired by the local municipality to clean up public parks. Three different combinations of labor and physical capital for cleaning up a single average-sized park appear in Table 1. The first production technology is heavy on workers and light on machines, while the next two technologies substitute machines for workers. Since all three of these production methods produce the same thing—one

cleaned-up park—a profit-seeking firm will choose the production technology that is least expensive, given the prices of labor and machines.

Production technology 1	10 workers	2 machines
Production technology 2	7 workers	4 machines
Production technology 3	3 workers	7 machines

Three Ways to Clean a Park

Production technology 1 uses the most labor and least machinery, while production technology 3 uses the least labor and the most machinery. [link] outlines three examples of how the total cost will change with each production technology as the cost of labor changes. As the cost of labor rises from example A to B to C, the firm will choose to substitute away from labor and use more machinery.

Example A: Workers cost R40/hr, machines cost R80/hr			
	Labor Cost/hr	Machine Cost/hr	Total Cost/hr
Cost of technology 1	10 × R40 = R400	2 × R80 = R160	R560
Cost of technology 2	7 × R40 = R280	4 × R80 = R320	R600

Cost of technology 3	3 × R40 = R120	7 × R80 = R560	R680
Example B: Workers cost R55/hr, machines cost R80/hr			
	Labor Cost/hr	Machine Cost/hr	Total Cost/hr
Cost of technology 1	10 × R55 = R550	2 × R80 = R160	R710
Cost of technology 2	7 × R55 = R385	4 × R80 = R320	R705
Cost of technology 3	3 × R55 = R165	7 × R80 = R560	R725
Example C: Workers cost R90/hr, machines cost R80/hr			
	Labor Cost/hr	Machine Cost/hr	Total Cost/hr
Cost of technology 1	10 × R90 = R900	2 × R80 = R160	R1,060
Cost of technology 2	7 × R90 = R630	4 × R80 = R320	R950
Cost of technology 3	3 × R90 = R270	7 × R80 = R560	R830

Total Cost with Rising Labor Costs

Example A shows the firm’s cost calculation when wages are R40 per hour and machines costs are R80 per hour. In this case, technology 1 is the low-cost production technology. In example B, wages rise to R55 per hour,

while the cost of machines does not change, in which case technology 2 is the low-cost production technology. If wages keep rising up to R90 per hour, while the cost of machines remains unchanged, then technology 3 clearly becomes the low-cost form of production, as shown in example C.

This example shows that as an input becomes more expensive (in this case, the labor input), firms will attempt to conserve on using that input and will instead shift to other inputs that are relatively less expensive. This pattern helps to explain why the demand curve for labor (or any input) slopes down; that is, as labor becomes relatively more expensive, profit-seeking firms will seek to substitute the use of other inputs. When a multinational employer like Coca-Cola or McDonald's sets up a bottling plant or a restaurant in a high-wage economy like the United States, Canada, Japan, or Western Europe, it is likely to use production technologies that conserve on the number of workers and focuses more on machines. However, that same employer is likely to use production technologies with more workers and less machinery when producing in a relatively lower-wage country like South Africa, Mexico or India.

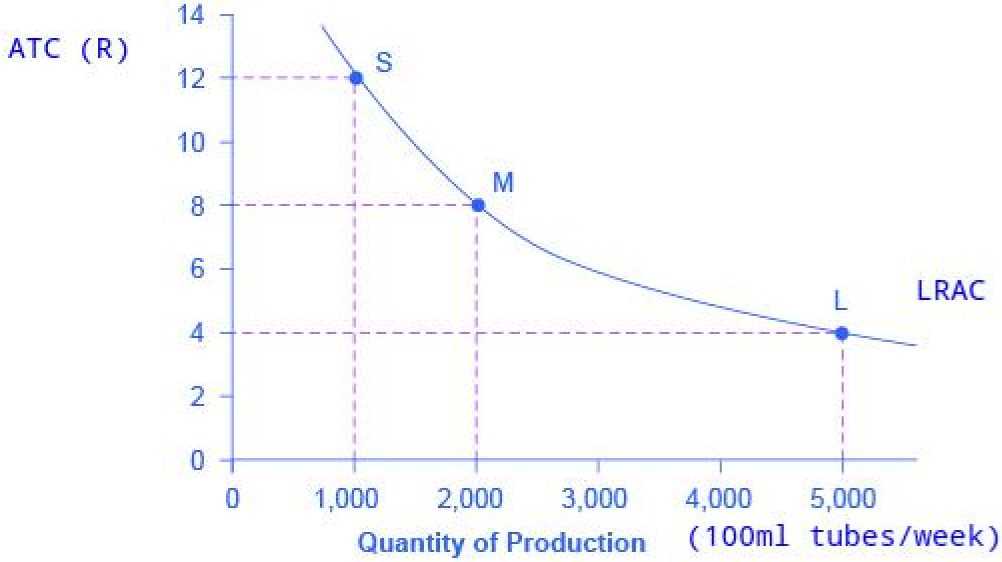
Economies of Scale

Once a firm has determined the least costly production technology, it can consider the optimal scale of production, or quantity of output to produce. Many industries experience economies of scale. Economies of scale refers to the situation where, as the quantity of output goes up, the cost per unit (average cost) goes down. This is the idea behind "warehouse stores" like Makro or Game. In everyday language: a larger factory can produce at a lower average cost than a smaller factory.

Figure 1 illustrates the idea of economies of scale, showing the average cost of producing 100 ml tubes of no-name-brand toothpaste as the quantity of output rises. For a small-sized factory like S, with an output level of 1,000 tubes per week, the average cost of production is R12 per 100 ml tube. For a medium-sized factory like M, with an output level of 2,000 tubes per week, the average cost of production falls to R8 per tube. For a large factory like L, with an output of 5,000 tubes per week, the average cost of production declines still further to R4 per 100 ml tube. Again, please note

that the cost of producing a tube of toothpaste is not necessarily the price at which it will eventually be sold. Cost and price are different concepts and should not be confused.

Economies of Scale



A small factory like S produces 1,000 tubes per week of toothpaste at an average cost of R12 per tube. A medium factory like M produces 2,000 tubes per week at a cost of R8 per tube. A large factory like L produces 5,000 tubes per week at a cost of R4 per tube. Economies of scale exist because the larger scale of production leads to lower average costs.

The average cost curve in Figure 1 may appear similar to the **short-run average cost curves** presented earlier in this chapter, although it is downward-sloping rather than U-shaped. But there is one major difference. The economies of scale curve is a **long-run average cost curve** (LRAC), because it allows all factors of production to change. The short-run average cost curves (SRAC) presented earlier in this chapter assumed the existence of fixed costs, and only variable costs were allowed to change.

One prominent example of economies of scale occurs in the chemical industry. Chemical plants such as SASOL have a lot of pipes. The cost of the materials for producing a pipe is related to the circumference of the pipe

and its length. However, the volume of chemicals that can flow through a pipe is determined by the cross-section area of the pipe. The calculations in Table 3 show that a pipe which uses twice as much material to make (as shown by the circumference of the pipe doubling) can actually carry four times the volume of chemicals because the cross-section area of the pipe rises by a factor of four (as shown in the Area column). Pipe sizes are often measured in inches (1 inch = 2.54 cm)

	Circumference ($2\pi r$)	**Area (πr^2)**
4-inch pipe	12.5 inches	12.5 square inches
8-inch pipe	25.1 inches	50.2 square inches
16-inch pipe	50.2 inches	201.1 square inches

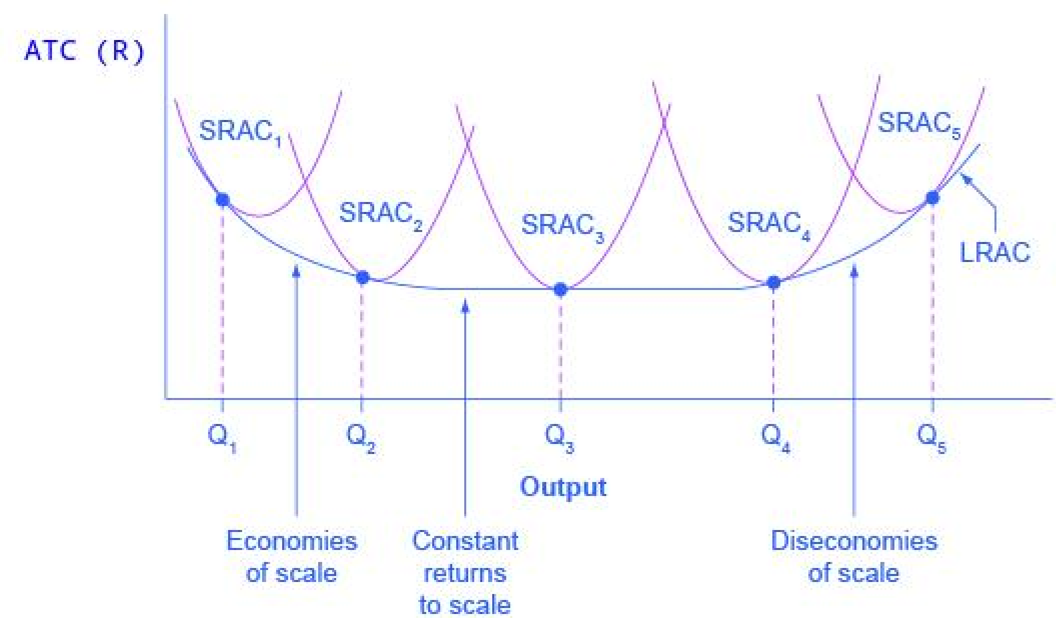
Comparing Pipes: Economies of Scale in the Chemical Industry

A doubling of the cost of producing the pipe allows the chemical firm to process four times as much material. This pattern is a major reason for economies of scale in chemical production, which uses a large quantity of pipes. It is quite logical that the cost per unit of chemicals must fall if the total cost is being divided by four times as much chemicals. Of course, economies of scale in a chemical plant are more complex than this simple calculation suggests. But the chemical engineers who design these plants have long used what they call the “six-tenths rule,” a rule of thumb which holds that increasing the quantity produced in a chemical plant by a certain percentage will increase total cost by only six-tenths (or 60%) as much.

Shapes of Long-Run Average Cost Curves

While in the short run firms are limited to operating on a single average cost curve (corresponding to the level of fixed costs they have chosen), in the long run when all costs are variable, they can choose to operate on any average cost curve. Thus, the **long-run average cost (LRAC) curve** is actually based on a group of **short-run average cost (SRAC) curves**, each of which represents one specific level of fixed costs. More precisely, the long-run average cost curve will be the least expensive average cost curve for any level of output. Figure 2 shows how the long-run average cost curve is built from a group of short-run average cost curves. Five short-run-average cost curves appear on the diagram. Each SRAC curve represents a different level of fixed costs. For example, you can imagine $SRAC_1$ as a small factory, $SRAC_2$ as a medium factory, $SRAC_3$ as a large factory, and $SRAC_4$ and $SRAC_5$ as very large and ultra-large. Although this diagram shows only five SRAC curves, presumably there are an infinite number of other SRAC curves between the ones that are shown. This family of short-run average cost curves can be thought of as representing different choices for a firm that is planning its level of investment in fixed cost physical capital—knowing that different choices about capital investment in the present will cause it to end up with different short-run average cost curves in the future.

From Short-Run Average Cost Curves to Long-Run Average Cost Curves



The five different short-run average cost (SRAC) curves each

represents a different level of fixed costs, from the low level of fixed costs at $SRAC_1$ to the high level of fixed costs at $SRAC_5$. Other SRAC curves, not shown in the diagram, lie between the ones that are shown here. The long-run average cost (LRAC) curve shows the lowest cost for producing each quantity of output when fixed costs can vary, and so it is formed by the bottom edge of the family of SRAC curves. If a firm wished to produce quantity Q_3, it would choose the fixed costs associated with $SRAC_3$.

The long-run average cost curve shows the cost of producing each quantity in the long run, when the firm can choose its level of fixed costs and thus choose which short-run average costs it desires. If the firm plans to produce in the long run at an output of Q_3, it should make the set of investments that will lead it to locate on $SRAC_3$, which allows producing Q_3 at the lowest cost. A firm that intends to produce Q_3 would be foolish to choose the level of fixed costs at $SRAC_2$ or $SRAC_4$. At $SRAC_2$ the level of fixed costs is too low for producing Q_3 at lowest possible cost, and producing Q_3 would require adding a very high level of variable costs and make the average cost very high. At $SRAC_4$, the level of fixed costs is too high for producing q_3 at lowest possible cost, and again average costs would be very high as a result.

The shape of the long-run cost curve, as drawn in Figure 2, is fairly common for many industries. The left-hand portion of the long-run average cost curve, where it is downward- sloping from output levels Q_1 to Q_2 to Q_3, illustrates the case of **economies of scale** . In this portion of the long-run average cost curve, larger scale leads to lower average costs. This pattern was illustrated earlier in Figure 1.

In the middle portion of the long-run average cost curve, the flat portion of the curve around Q_3, economies of scale have been exhausted. In this situation, allowing all inputs to expand does not much change the average cost of production, and it is called **constant returns to scale**. In this range of the LRAC curve, the average cost of production does not change much as scale rises or falls. The right hand portion of the LRAC (upward sloping) illustrates **diseconomies of scale** from output levels Q_4 to Q_5. In this

portion of the LRAC, larger scale leads to higher unit or average costs of production. So what are the reasons that LRAC has this u-shape?

Underlying reasons for economies and diseconomies of scale
The bigger firms get, the more easily they are able to hire specialist labour which tends to be more efficient than generalist labor employed by smaller firms. Similarly, bigger firms are able to employ specialist managers who, because they focus only on one task, are able to operate more efficiently than a generalist manager of a smaller firm who has to perform all the management tasks (sales, marketing, finance, human resources etc). Bigger firms are also able to purchase more efficient capital (machines and equipment) that are capable of processing larger volumes of work more quickly than the smaller machines and equipment that smaller firms can afford. Smaller firms cannot secure credit at the same preferential rate enjoyed by larger firms that have more collateral to offer credit providers such as banks and suppliers. Larger firms also tend to have product brands that are more recognisable than those of smaller firms so proportionately they do not have to spend as much as smaller firms on advertising and marketing their products (marketing expenditure as a percent of sales turnover).

The main factor that causes diseconomies of scale is the difficulty of managing a firm's operations as it becomes a large-scale producer. Contrary to this, in a small firm the owner is close to the production line, understands the firm's operations, can keep in touch with each member of his workforce and so is able to process information and make informed decisions. In large firms with different divisions and many levels of management communication and cooperation become more difficult. Decision making is slowed down as it takes longer for information to filter through to the different management levels. Employees may become disgruntled and frustrated and this may affect their work ethic and productivity and so the unit (average) cost of production will start to increase: diseconomies of scale. Firms that cut back their operations are often responding to finding themselves in the diseconomies region, thus moving back to a lower average cost at a lower output level. There are many examples of firms rationalising their operations in order to reduce costs and improve

profitability. Is the concept of diseconomies of scale related to that of diminishing returns?

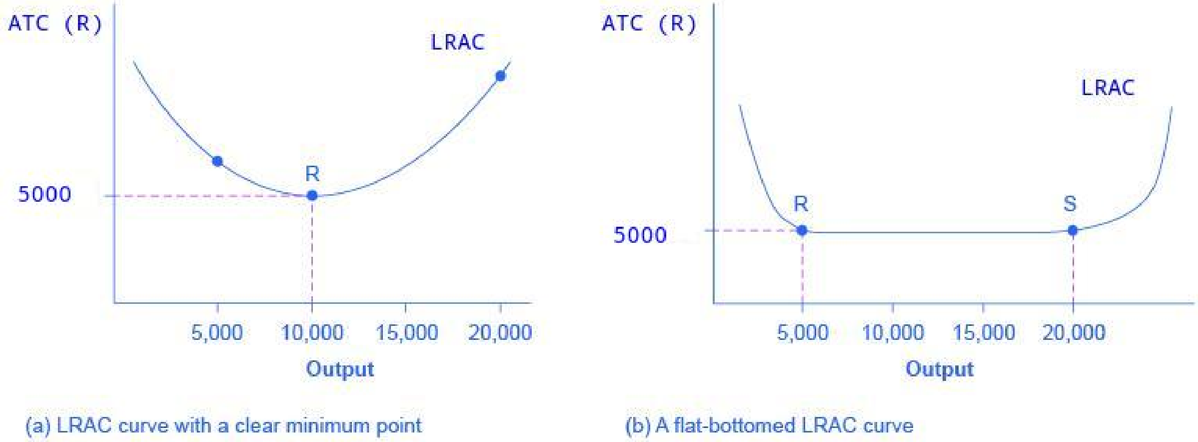
Note:
How do economies of scale compare to diminishing marginal returns?
The concept of economies of scale, where average costs decline as production expands, might seem to conflict with the idea of diminishing marginal returns, where marginal costs rise as production expands. But diminishing marginal returns refers only to the short-run average cost curve, where one variable input (like labor) is increasing, but other inputs (like capital) are fixed. Economies of scale refers to the long-run average cost curve where all inputs are being allowed to increase together. Thus, it is quite possible and common to have an industry that has both diminishing marginal returns when only one input is allowed to change, and at the same time has increasing or constant economies of scale when all inputs change together to produce a larger-scale operation.

The Size and Number of Firms in an Industry

The shape of the long-run average cost curve has implications for how many firms will compete in an industry, and whether the firms in an industry have many different sizes, or tend to be the same size. For example, say that one million dishwashers are produced every year at a unit cost of R5000 each and the long-run average cost curve for dishwashers is shown in the diagram below. In (a), the lowest point of the LRAC curve occurs at a quantity of 10,000 dishwashers produced for a typical firm. Thus, the market for dishwashers could consist of 100 different manufacturing plants of roughly the same size. If, however, some firms build factories that produced 5,000 dishwashers per year or 25,000 dishwashers per year as shown in (b), the average costs of production at such plants would be well above R5000 per unit, and the firms would not be able to compete. They are either too small or too big to be competitive. So in this particular industry, the nature of LRAC dictates a factory or plant size with the capacity to produce between 5,000 - 20,000 units in order to

achieve the lowest possible unit cost of manufacturing a dishwasher. A firm that wants to compete effectively in the dishwasher market must therefore be able to produce between 5,000 - 20,000 units.

The LRAC Curve and the Size and Number of Firms



(a) Low-cost firms will produce at output level R. When the LRAC curve has a clear minimum point, then any firm producing a different quantity will have higher costs. In this case, a firm producing at a quantity of 10,000 will produce at a lower average cost than a firm producing, say, 5,000 or 20,000 units. (b) Low-cost firms will produce between output levels R and S. When the LRAC curve has a flat bottom, then firms producing at any quantity along this flat bottom can compete. In this case, any firm producing a quantity between 5,000 and 20,000 can compete effectively, although firms producing less than 5,000 or more than 20,000 would face higher average costs and be unable to compete.

Note:

How can cities be viewed as examples of economies of scale?

Why are people and economic activity concentrated in cities, rather than distributed evenly across a country? The fundamental reason must be related to the idea of economies of scale—that grouping economic activity is more productive in many cases than spreading it out. For example, cities

provide a large group of nearby customers, so that businesses can produce at an efficient economy of scale. They also provide a large group of workers and suppliers, so that business can hire easily and purchase whatever specialized inputs they need. Many of the attractions of cities, like sports stadiums and museums, can operate only if they can draw on a large nearby population base. Cities are big enough to offer a wide variety of products, which is what many shoppers are looking for.

These factors are not exactly economies of scale in the narrow sense of the production process of a single firm, but they are related to growth in the overall size of population and market in an area. Cities are sometimes called "agglomeration economies."

These agglomeration factors help to explain why every economy, as it develops, has an increasing proportion of its population living in urban areas. In 2010, Johannesburg was the 53rd largest city in the world (The largest cities in the world and their mayors: 2011). It has been estimated that, by 2050, about 80% of South Africa's population will be living in urban areas (SA's urban population growing larger and younger: 2015). In much of Africa, the proportion of the population in urban areas is only about 30%. One of the great challenges for these countries as their economies grow will be to manage the growth of the great cities that will arise.

If cities offer economic advantages that are a form of economies of scale, then why don't all or most people live in one giant city? At some point, agglomeration economies must turn into diseconomies. For example, traffic congestion may reach a point where the gains from being geographically nearby are counterbalanced by how long it takes to travel. Road traffic congestion in Johannesburg, Pretoria and Cape Town, for example, can mean delays of up to two hours to travel 10km! High densities of people, cars, and factories can mean more rubbish and air and water pollution. Facilities like parks or museums may become overcrowded. There may be economies of scale for negative activities like crime, because high densities of people and businesses, combined with the greater impersonality of cities, make it easier for illegal activities as well as legal ones. The future of cities, both in South Africa and in other countries around the world, will be determined by their ability to benefit from the economies of agglomeration and to minimize or counterbalance the corresponding diseconomies.

Thus, the shape of the long-run average cost curve reveals whether competitors in the market will be different sizes. If the LRAC curve has a single point at the bottom as in Figure 3(a), then the firms in the market will be about the same size, but if the LRAC curve has a flat-bottomed segment of constant returns to scale as in Figure 3(b), then firms in the market may be a variety of different sizes. So in our dishwasher manufacturer example, firms producing between 5,000 and 20,000 dishwashers will all be able to compete in this market. In Figure 3(a), however, only firms that are capable of producing about 10,000 dishwashers will be able to compete effectively.

The relationship between the quantity at the minimum of the long-run average cost curve and the quantity demanded in the market at that price will predict how much competition is likely to exist in the market. If the quantity demanded in the market far exceeds the quantity at the minimum of the LRAC, then many firms will compete. If the quantity demanded in the market is only slightly higher than the quantity at the minimum of the LRAC, a few firms will compete. If the quantity demanded in the market is less than the quantity at the minimum of the LRAC, a single-producer monopoly is a likely outcome.

Shifting Patterns of Long-Run Average Cost

New developments in production technology can shift the long-run average cost curve in ways that can alter the size distribution of firms in an industry.

For much of the twentieth century, the most common change has been to see alterations in technology, like the assembly line or the large department store, where large-scale producers seemed to gain an advantage over smaller ones. In the long-run average cost curve, the downward-sloping economies of scale portion of the curve stretched over a larger quantity of output.

However, new production technologies do not inevitably lead to a greater average size for firms. For example, in recent years some new technologies for generating electricity on a smaller scale have appeared. The traditional coal-burning electricity plants needed to produce 300 to 600 megawatts of power to exploit economies of scale fully. However, high-efficiency

turbines to produce electricity from burning natural gas can produce electricity at a competitive price while producing a smaller quantity of 100 megawatts or less. These new technologies create the possibility for smaller companies or plants to generate electricity as efficiently as large ones. Another example of a technology-driven shift to smaller plants may be taking place in the tyre industry. A traditional mid-size tyre plant produces about six million tires per year. However, in 2000, the Italian company Pirelli introduced a new tyre factory that uses many robots. The Pirelli tyre plant produced only about one million tyres per year, but did so at a lower average cost than a traditional mid-sized tyre plant.

There has been much controversy in recent years over whether the new information and communications technologies will lead to a larger or smaller size for firms. On one hand, the new technology may make it easier for small firms to reach out beyond their local geographic area and find customers in other cities, provinces or even countries. This factor might seem to predict a future with a larger number of small competitors. On the other hand, perhaps the new information and communications technology will create "winner-take-all" markets where one or a few large companies will tend to command a large share of total sales as have Vodacom, MTN and Cell C in South Africa's mobile phone market. Or as Microsoft has done in the production of software for personal computers or Amazon has done in online bookselling. Moreover, improved information and communication technologies might make it easier to manage many different plants and operations across the country or around the world, and thus encourage larger firms. This ongoing battle between the forces of smallness and largeness will be of great interest to economists, business people, and policymakers.

Note:
Amazon
Traditionally, bookstores have operated in retail locations with inventories held either on the shelves or in the back of the store. These retail locations were very pricey in terms of rent. Amazon has no retail locations; it sells online and delivers by mail. Amazon offers almost any book in print,

convenient purchasing, and prompt delivery by mail. Amazon holds its inventories in huge warehouses in low-rent locations around the world. The warehouses are highly computerized using robots and relatively low-skilled workers, making for low average costs per sale. Amazon demonstrates the significant advantages economies of scale can offer to a firm that exploits those economies.

Key Concepts and Summary

A production technology refers to a specific combination of labor, physical capital, and technology that makes up a particular method of production.

In the long run, firms can choose their production technology, and so all costs become variable costs. In making this choice, firms will try to substitute relatively inexpensive inputs for relatively expensive inputs where possible, so as to produce at the lowest possible long-run average cost.

Economies of scale refers to a situation where as the level of output increases, the average cost decreases. Constant returns of scale refers to a situation where average cost does not change as output increases. Diseconomies of scale refers to a situation where as output increases, average costs increase also.

The long-run average cost curve shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology. A downward-sloping LRAC shows economies of scale; a flat LRAC shows constant returns of scale; an upward-sloping LRAC shows diseconomies of scale. If the long-run average cost curve has only one quantity produced that results in the lowest possible average cost, then all of the firms competing in an industry should be the same size. However, if the LRAC has a flat segment at the bottom, so that a range of different quantities can be produced at the lowest average cost, the firms competing in the industry will display a range of sizes. The market demand in conjunction with the long-run average cost curve determines how many firms will exist in a given industry.

If the quantity demanded in the market of a certain product is much greater than the quantity found at the bottom of the long-run average cost curve, where the cost of production is lowest, the market will have many firms competing. If the quantity demanded in the market is less than the quantity at the bottom of the LRAC, there will likely be only one firm.

Self-Check Questions

Exercise:

Problem:

Return to the problem explained in Table 1 and Table 2. If the cost of labor remains at R40 per hour, but the cost of a machine decreases to R50 per hour, what would be the total cost of each method of production? Which method should the firm use, and why?

Solution:

The new table should look like this:

The firm should choose production technology 3 since it has the lowest total cost. This makes sense since, with cheaper machine hours, one would expect a shift in the direction of more machines and less labor.

Table 3: Exercise 1 Solution Table

	Labor cost (R/hr)	Machine cost (R/hr)	Total cost (R/hr)
METHOD 1	10 x R40	2 x R50	R 500.00
METHOD 2	7 x R40	4 x R50	R 480.00
METHOD 3	3 x R40	7 x R50	R 470.00

Exercise:

Problem:

Suppose the cost of machines increases to R55 per hour, while the cost of labor stays at R40 per hour. How would that affect the total cost of the three methods? Which method should the firm choose now?

Solution:

The firm should choose production technology 2 since it has the lowest total cost. Because the cost of machines increased (relative to the previous question), you would expect a shift toward less capital and more labor.

Table 4: Exercise 2 Solution Table

	Labor cost (R/hr)	Machine cost (R/hr)	Total cost (R/hr)
METHOD 1	10 x R40	2 x R55	R 510.00
METHOD 2	7 x R40	4 x R55	R 500.00
METHOD 3	3 x R40	7 x R55	R 505.00

Exercise:

Problem:

Car manufacturing is an industry subject to significant economies of scale. Suppose there are four domestic car manufacturers, but the demand for domestic cars is no more than 2.5 times the quantity produced at the bottom of the long-run average cost curve. What do you expect will happen to the domestic car industry in the long run?

Solution:

This might be close to what the situation in South Africa is with the view that local car manufacturers are producing too many makes of cars for the available demand (Hardigree, 2014). Since there is only

demand enough for 2.5 firms to reach the bottom of the average cost curve, you would expect one firm will not be around in the long run, and at least one firm will be struggling.

Review Questions

Exercise:

Problem:

What shapes would you generally expect each of the following cost curves to have: fixed costs, variable costs, marginal costs, average total costs, and average variable costs?

Exercise:

Problem: What is a production technology?

Exercise:

Problem:

In choosing a production technology, how will firms react if one input becomes relatively more expensive?

Exercise:

Problem: What is a long-run average cost curve?

Exercise:

Problem:

What is the difference between economies of scale, constant returns of scale, and diseconomies of scale?

Exercise:

Problem:

What shape of a long-run average cost curve illustrates economies of scale, constant returns to scale, and diseconomies of scale?

Exercise:

Problem:

Why will firms in most markets be located at or close to the bottom of the long-run average cost curve?

Critical Thinking Questions

Exercise:

Problem:

It is clear that businesses operate in the short run, but do they ever operate in the long run? Discuss.

Exercise:

Problem:

How would an improvement in technology, like the high-efficiency gas turbines, or, Pirelli tyre plant robots, affect the long-run average cost curve of a firm? Can you draw the old curve and the new one on the same axes? How might such an improvement affect other firms in the industry?

Exercise:

Problem:

Do you think that the minibus/kombi taxi industry in large cities such as Durban or Johannesburg would be subject to significant economies of scale? Why or why not?

Problems

Exercise:

Problem:

A small Durban BBBEE compliant company has been awarded a tender by the eThekwini Municipality to keep Durban's streets clean. It has been assigned 100 streets to sweep and keep clear of litter. It can use various combinations of capital and labor: lots of labor with brooms and rakes and just a few mechanical brooms, less labor with hand-held mechanical blowers and industrial vacuum cleaners, and still less labor with a pickup truck that has a mechanical broom and vacuum cleaner attached to the front of the vehicle. To summarize, the possible production/operation methods are:

Method 1: 50 units of labor, 10 units of capital

Method 2: 20 units of labor, 40 units of capital

Method 3: 10 units of labor, 70 units of capital

If hiring labor costs R30 per unit and a unit of capital costs R400, what production method should be chosen? What method should be chosen if the cost of labor rises to R60 per unit?

References

Hardigree, M. 2014. Carmakers starting to realize they make too many damn models. Jalopnik, 14 Nov. Available http://jalopnik.com/this-is-the-morning-shift-our-one-stop-daily-roundup-o-1658771171 (Accessed: 25 Apr 2016)

The largest cities in the world and their mayors: cities ranked 1 to 150. 2011. City Mayors Statistics. Available http://www.citymayors.com/statistics/largest-cities-mayors-1.html (Accessed: 25 Apr 2016)

SA’s urban population ‘growing larger and younger’. 2015. Business Day Live, 26 May. http://www.bdlive.co.za/national/2015/05/26/sas-urban-population-growing-larger-and-younger (Accessed: 25 Apr 2016)

Glossary

constant returns to scale
expanding all inputs proportionately does not change the average cost of production

diseconomies of scale
the long-run average cost of producing each individual unit increases as total output increases

long-run average cost (LRAC) curve
shows the lowest possible average cost of production, allowing all the inputs to production to vary so that the firm is choosing its production technology

production technologies
alternative methods of combining inputs to produce output

short-run average cost (SRAC) curve
the average total cost curve in the short term; shows the total of the average fixed costs and the average variable costs

Perfect Competition and Why It Matters
By the end of this section, you will be able to:

- Explain the characteristics of a perfectly competitive market
- Discuss how perfectly competitive firms react in the short run and in the long run

Firms are said to be in **perfect competition** when the following conditions occur: (1) many firms produce identical products; (2) many buyers are available to buy the product, and many sellers are available to sell the product; (3) sellers and buyers have all relevant information to make rational decisions about the product being bought and sold; and (4) firms can enter and leave the market without any restrictions—in other words, there is free entry and exit into and out of the market.

A perfectly competitive firm is known as a **price taker**, because the pressure of competing firms forces them to accept the current equilibrium price in the market. If a firm in a perfectly competitive market raises the price of its product by so much as a cent, it will lose all of its sales to competitors. In such a market if a maize farmer, for example, wants to know what the going price of maize is, he or she has to check the commodity prices online (internet) or the financial news on TV or announced on the radio. The market price is determined solely by supply and demand in the entire market and not the individual farmer. Also, a perfectly competitive firm must be a very small player in the overall market, so that it can increase or decrease output without noticeably affecting the overall quantity supplied and price in the market.

A perfectly competitive market is not a realistic situation since it is an extreme case. However, in real life producers in a number of industries do face many competitor firms selling highly similar goods, in which case they must often act as price takers. Agricultural markets are often used as an example. The same crops grown by different farmers are largely interchangeable. Another example of a highly competitive market is that of motor vehicles, especially on a global scale. Other examples include the markets for computer software or applications (apps such as games, social media, personal finance apps and many more). In fact this market is so competitive that many software developers simply **give** part of what they

produce away as an incentive to customers to buy other products from them. The publishing and music industries are similarly competitive and also give away a certain amount of their products to entice and attract customers. As an interesting exercise do an internet search to see how many free apps, music, e-books,and online magazine and newspaper articles can be downloaded.

Note:
Visit this website that reveals the current value of various commodities.

This chapter examines how profit-seeking firms decide how much to produce in perfectly competitive markets. In the short run, the perfectly competitive firm will seek the quantity of output where profits are highest or, if profits are not possible, where losses are lowest. In this example, the "short run" refers to a situation in which firms are producing with one fixed input and incur fixed costs of production. (In the real world, firms can have many fixed inputs.)

In the long run, perfectly competitive firms will react to profits by increasing production. They will respond to losses by reducing production or exiting the market. Ultimately, a long-run *equilibrium* will be attained when no new firms want to enter the market and existing firms do not want to leave the market, as economic profits have been driven down to zero.

Key Concepts and Summary

A perfectly competitive firm is a price taker, which means that it must accept the equilibrium price at which it sells goods. If a perfectly competitive firm attempts to charge even a tiny amount more than the market price, it will be unable to make any sales. In a perfectly competitive market there are thousands of sellers, easy entry, and identical products.

Firms in perfectly competitive markets will make different choices depending on whether it is the **"short-run"** or the **"long-run"** production period. Generally, the long-run is associated with greater flexibility for firms than they would normally enjoy in the short-run. A short-run production period is when firms are producing with some fixed inputs (at least one fixed input). The long-run refers to a production period in which all inputs are variable. Long-run equilibrium in a perfectly competitive industry occurs after all firms have entered and exited the industry (and firms have this flexibility only in the long run) and seller profits are driven to zero. We will observe this process in the following sections.

Perfect competition means that there are many sellers, there is easy entry and exiting of firms, products are identical from one seller to another, and sellers are price takers.

Self-Check Questions

Exercise:

Problem:

Firms in a perfectly competitive market are said to be “price takers”—that is, once the market determines an equilibrium price for the product, firms must accept this price. If you sell a product in a perfectly competitive market, but you are not happy with its price, would you raise the price, even by a cent?

Solution:

No, you would not raise the price. Your product is exactly the same as the product of the many other firms in the market. If your price is greater than that of your competitors, then your customers would

switch to them and stop buying from you. You would lose all your sales.

Exercise:

Problem:

Would independent trucking fit the characteristics of a perfectly competitive industry?

Solution:

Possibly. Independent truckers are by definition small and numerous. All that is required to get into the business is a truck (not an inexpensive asset, though) and a commercial driver's license. To exit, one need only sell the truck. All trucks are essentially the same, providing transportation from point A to point B. (We're assuming we not talking about specialized trucks.) Independent truckers must take the going rate for their service, so independent trucking does seem to have most of the characteristics of perfect competition.

Review Questions

Exercise:

Problem:

A single firm in a perfectly competitive market is relatively small compared to the rest of the market. What does this mean? How "small" is "small"?

Exercise:

Problem:

What are the four basic assumptions of perfect competition? Explain in words what they mean for a perfectly competitive firm.

Exercise:

Problem: What is a “price taker” firm?

Critical Thinking Questions

Exercise:

Problem:

Finding a life partner is a complicated process that may take many years. It is hard to think of this process as being part of a very complex market, with a demand and a supply for partners. Think about how this market works and some of its characteristics, such as search costs. Would you consider it a perfectly competitive market?

Exercise:

Problem:

Can you name five examples of perfectly competitive markets? Why or why not?

Glossary

market structure
: the conditions in an industry, such as number of sellers, how easy or difficult it is for a new firm to enter, and the type of products that are sold

perfect competition
: each firm faces many competitors that sell identical products

price taker
: a firm in a perfectly competitive market that must take the prevailing market price as given

How Perfectly Competitive Firms Make Output Decisions

By the end of this section, you will be able to:

- Calculate profits by comparing total revenue and total cost
- Identify profits and losses with the average cost curve
- Explain the shutdown point
- Determine the price at which a firm should continue producing in the short run

A perfectly competitive firm has only one major decision to make—namely, what quantity to produce. To understand why this is so, consider a different way of writing out the basic definition of profit:

Equation:

$$\begin{aligned}\text{Profit} &= \text{Total revenue} - \text{Total cost}\\ &= (\text{Price})(\text{Quantity produced}) - (\text{Average cost})(\text{Quantity produced})\end{aligned}$$

Since a perfectly competitive firm must accept the price for its output as determined by the product’s market demand and supply, it cannot choose the price it charges. This is already determined in the profit equation, and so the perfectly competitive firm can sell any number of units at exactly the same price. It implies that the firm faces a perfectly elastic demand curve for its product: buyers are willing to buy any number of units of output from the firm at the market price. When the perfectly competitive firm chooses what quantity to produce, then this quantity—along with the prices prevailing in the market for output and inputs—will determine the firm’s total revenue, total costs, and ultimately, level of profits.

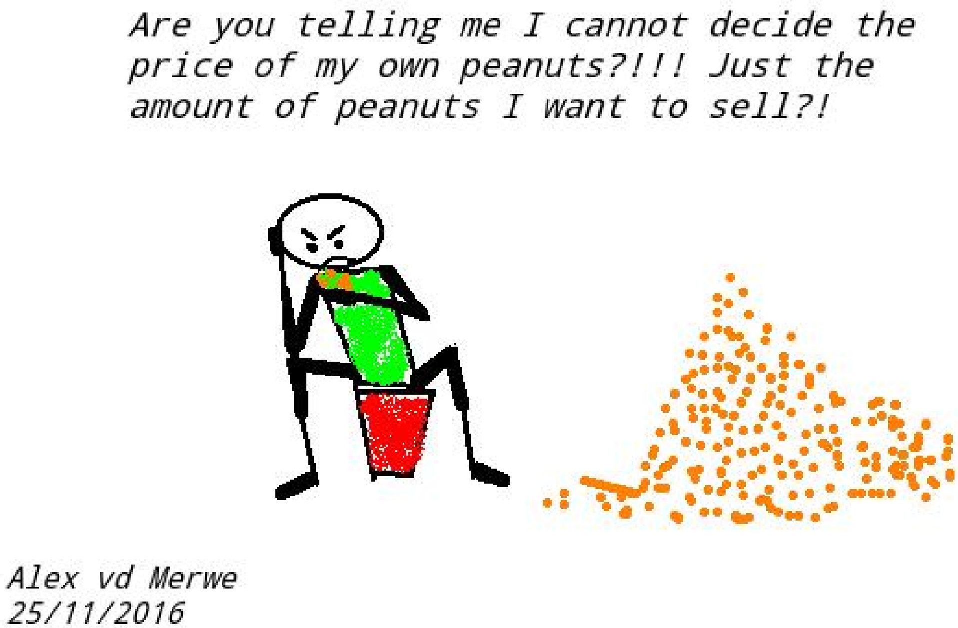


Figure 1

Determining the Highest Profit by Comparing Total Revenue and Total Cost

A perfectly competitive firm can sell as large a quantity as it wishes, as long as it accepts the current market price. Total revenue is going to increase as the firm sells more, depending on the price of the product and the number of units sold. If you increase the number of units sold at a given price, then total revenue will increase. If the price of the product increases for every unit sold, then total revenue also increases. As an example of how a perfectly competitive firm decides what quantity to produce, consider the case of a small-scale farmer in the uMkhanyakude district of Zululand who, along with many other small scale emerging farmers in the district, produces peanuts and sells them for R4 per 50 gram pack. Sales of one pack of peanuts will bring in R4, two packs will be R8, three packs will be R12, and so on. If, for example, the price of 50 gram packs of peanuts doubles to R8 per pack, then sales of one pack of peanuts will be R8, two packs will be R16, three packs will be R24, and so on.

Total revenue and total costs for the peanut farm, broken down into fixed and variable costs, are shown in Table 1 and also appear in Figure 2. The horizontal axis shows the quantity of peanuts produced in packs; the vertical axis shows both total revenue and total costs, measured in Rands. The total cost curve intersects with the vertical axis at a value that shows the level of fixed costs, and then slopes upward. All these cost curves follow the same characteristics as the curves covered in the Cost and Industry Structure chapter.

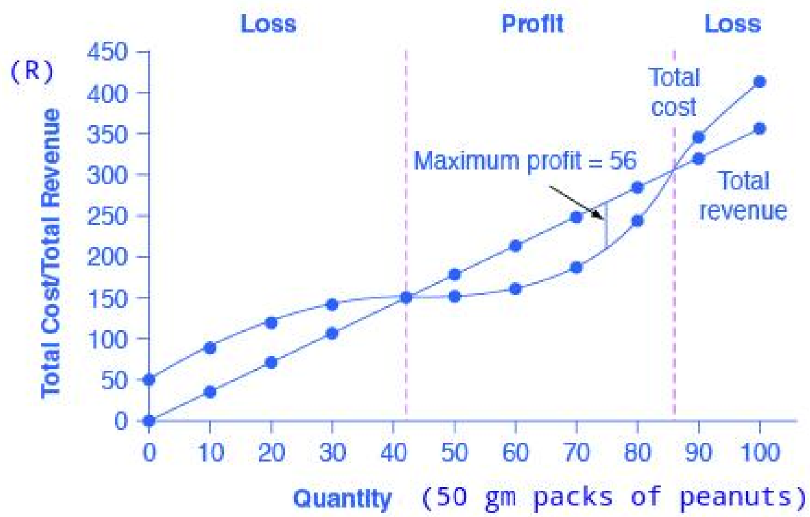


Figure 2: Total revenue for a perfectly competitive firm is a straight line sloping up. The slope is equal to the price of the good. Total cost also slopes up, but with some curvature. At higher levels of output, total cost begins to slope upward more steeply because of diminishing marginal returns. The maximum profit will occur at the quantity where the gap of total revenue over total cost is largest.

Table 1: Cost, Revenue and Profit at the Peanut Farm

Quantity, Q (50gm packs)	Total Cost , TC [R]	Fixed Cost, FC, [R]	Variable Cost, VC [R]	Total Revenue, TR [R]	Profit [R]
0	62	62	-	0	−62
10	90	62	28	40	−50
20	110	62	48	80	−30
30	126	62	64	120	−6
40	144	62	82	160	16
50	166	62	104	200	34
60	192	62	130	240	48
70	224	62	162	280	56
80	264	62	202	320	56
90	324	62	262	360	36
100	404	62	342	400	−4

Based on its total revenue and total cost curves, a perfectly competitive firm like the small-scale peanut farm can calculate the quantity of output that will provide the highest level of profit. At any given quantity, total revenue minus total cost will equal total profit. One way to determine the most profitable quantity to produce is to see at what quantity total revenue exceeds total cost by the largest amount. In Figure 2, the vertical gap between total revenue and total cost represents either profit (if total revenues are greater that total costs at a certain quantity) or losses (if total costs are greater that total revenues at a certain quantity). In this example, total costs will exceed total revenues at output levels from 0 to 40 packs, and so over this range of output, the firm will be making losses. At output levels from 50 to 80 packs, total revenues exceed total costs, so the firm is earning profits. But then at an output of 90 or 100 packs, total costs again exceed total revenues and the firm is making losses. Total profits appear in the final column of Table 1. The highest total profits in the table, as in the figure that is based on the table values, occur at an output of 70–80 packs, when profits will be R56.

A higher price would mean that total revenue would be higher for every quantity sold. A lower price would mean that total revenue would be lower for every quantity sold. What happens if the price drops low enough so that the total revenue line is completely below the total cost curve; that is, at every level of output, total costs are higher than total revenues? In this instance, the best the firm can do is to suffer losses. But a profit-maximizing firm will prefer the quantity of output where total revenues come closest to total costs and thus where the losses are smallest.

(Later we will see that sometimes it will make sense for the firm to shutdown, rather than stay in operation producing output.)

Comparing Marginal Revenue and Marginal Costs

Firms often do not have the necessary data they need to draw a complete total cost curve for all levels of production. They cannot be sure of what total costs would look like if they, say, doubled production or cut production in half, because they have not tried it. Instead, firms experiment. They produce a slightly greater or lower quantity and observe how profits are affected. In economic terms, this practical approach to maximizing profits means looking at how changes in production affect marginal revenue and marginal cost.

Figure 3 presents the marginal revenue and marginal cost curves based on the total revenue and total cost in Table 1. The **marginal revenue** curve shows the additional revenue gained from selling one more unit. As mentioned before, a firm in perfect competition faces a perfectly elastic demand curve for its product—that is, the firm's demand curve is a horizontal line drawn at the market price level. This also means that the firm's marginal revenue curve is the same as the firm's demand curve: Every time a consumer demands one more unit, the firm sells one more unit and revenue goes up by exactly the same amount equal to the market price. In this example, every time a 50gm pack of peanuts is sold, the firm's revenue increases by R4. Table 2 shows an example of this. This condition only holds for price taking firms in perfect competition where:
Equation:

$$\text{marginal revenue} = \text{price}$$

The formula for marginal revenue is:
Equation:

$$\text{marginal revenue} = \frac{\text{change in total revenue}}{\text{change in quantity}}$$

`Table 2: The Relationship Between Total and Marginal Revenue in Perfect Competition`

Price	Quantity	Total Revenue	Marginal Revenue
R4	1	R4	-
R4	2	R8	R4
R4	3	R12	R4
R4	4	R16	R4

Notice that marginal revenue does not change as the firm produces more output. That is because the price is determined by supply and demand and does not change as the farmer produces more (keeping in mind that, due to the relative small size of each firm, increasing their supply has no impact on the total market supply where price is determined).

Since a perfectly competitive firm is a price taker, it can sell whatever quantity it wishes at the market-determined price. Marginal cost, the cost per additional unit sold, is calculated by dividing the change in total cost by the change in quantity. The formula for marginal cost is:
Equation:

$$\text{marginal cost} = \frac{\text{change in total cost}}{\text{change in quantity}}$$

Ordinarily, marginal cost changes as the firm produces a greater quantity.

In the peanut farm example, marginal cost in Figure 3 (and also shown in Table 3) at first declines as production increases from 10 to 20 to 30 packs of peanuts—which represents the area of increasing marginal returns that is not uncommon at low levels of production. But then marginal costs start to increase, displaying the typical pattern of diminishing marginal returns. If the firm is producing at a quantity where MR is greater than MC, like 40 or 50 packs of peanuts, then it can increase profit by increasing output because the marginal revenue is exceeding the marginal cost. If the firm is producing at a quantity where MC is greater than MR, like 90 or 100 packs, then it can increase profit by reducing output because the reductions in marginal cost will exceed the reductions in marginal revenue. The firm's profit-maximizing choice of output will occur where MR = MC (or at a choice close to that point). This is the **profit maximising** (or loss minimising) rule that all firms try to follow. You will notice that what occurs on the production side is exemplified on the cost side. This is referred to as duality.

Market demand versus individual firm demand
Note that market supply and market demand determine the price of peanuts (R4 per pack) in a perfectly competitive market as shown in Figure 4. Note also the much larger quantity of peanuts traded (8000 packs). Each farmer in this competitive market will have a small share of this total trade. It is logical that market demand for peanuts will be more inelastic (downward sloping) than the perfectly elastic demand facing individual peanut farmers (the marginal revenue = average revenue = demand curve in Figure 3). This is because consumers have only one market (or maybe just a few markets) to choose from if they wish to buy peanuts compared to the many individual peanut farmers they could buy from. More choice, more substitutes, more elastic demand.

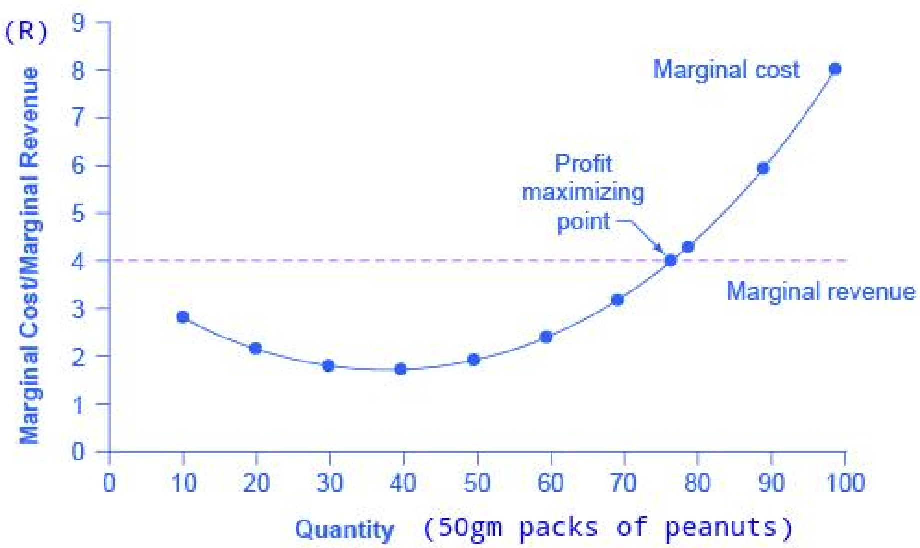


Figure 3: For a perfectly competitive firm, the marginal revenue (MR) curve is a horizontal straight line because it is equal to the price of the good, which is determined by the market, shown below. The marginal cost (MC) curve is sometimes first downward-sloping, if there is a region of increasing marginal returns at low levels of output, but is eventually upward-sloping at higher levels of output as diminishing marginal returns kick in.

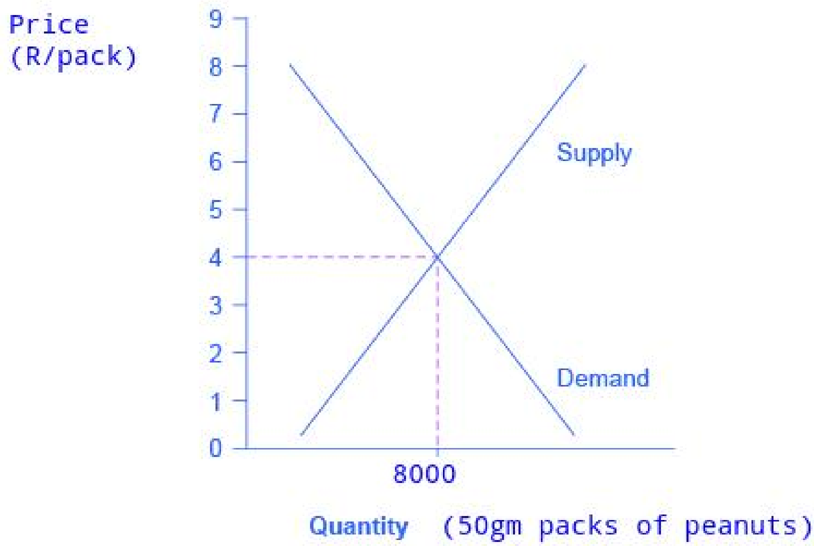


Figure 4: The equilibrium price of 50gm packs of peanuts is determined through the interaction of market supply and market demand at R4.00.

Table 3: Marginal Revenue and Marginal Cost at the Peanut Farm

Quantity (50gm packs)	Total Cost [R]	Fixed Cost [R]	Variable Cost [R]	Marginal Cost [R]	Total Revenue [R]	Marginal Revenue [R]
0	62	62	-	-	-	-
				2.8		
10	90	62	28		40	
				2		4
20	110	62	48		80	
				1.6		4
30	126	62	64		120	
				1.8		4
40	144	62	82		160	
				2.2		4
50	166	62	104		200	
				2.6		4
60	192	62	130		240	
				3.2		4
70	224	62	162		280	
				4		4
80	264	62	202		320	
				6		4
90	324	62	262		360	
				8		4
100	404	62	342		400	

In this example, the marginal revenue and marginal cost curves cross at a price of R4 and a quantity of 80 packs produced. If the farmer started out producing at a level of 60 packs, and then experimented with increasing production to 70 packs, marginal revenues from the increase in production would exceed marginal costs—and so profits would rise. The farmer has an incentive to keep producing. From a level of 70 to 80 packs, marginal cost and marginal revenue are equal so profit doesn’t change. If the farmer then experimented further with increasing production from 80 to 90 packs, he would find that marginal costs from the increase in production are greater than marginal revenues, and so profits would decline.

The profit-maximizing choice for a perfectly competitive firm will occur where marginal revenue is equal to marginal cost—that is, where MR = MC. A profit-seeking firm should keep expanding production as long as MR > MC. But at the level of output where MR = MC, the firm should recognize that it has achieved the highest possible level of economic profits. (In the example above, the profit maximizing output level is between 70 and 80 units of output, but the firm will not know they’ve maximized profit until they reach 80, where MR = MC.) Expanding

production into the zone where MR < MC will only reduce economic profits. Because the marginal revenue received by a perfectly competitive firm is equal to the price P, so that P = MR, the profit-maximizing rule for a perfectly competitive firm can also be written as a recommendation to produce at the quantity where P = MC.

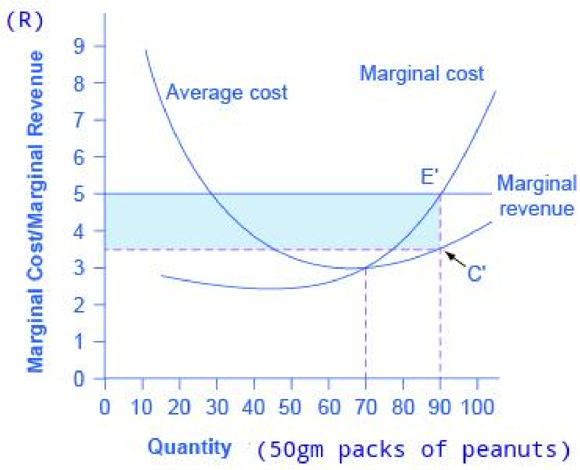
Profits and Losses with the Average Cost Curve

Does maximizing profit (producing where MR = MC) imply an actual economic profit? The answer depends on the relationship between price and average total cost. If the price that a firm charges is higher than its average cost of production for that quantity produced, then the firm will earn profits. Conversely, if the price that a firm charges is lower than its average cost of production, the firm will suffer losses. You might think that, in this situation, the farmer may want to shut down immediately. Remember, however, that the firm has already paid for fixed costs, such as equipment, so it may continue to produce and incur a loss.

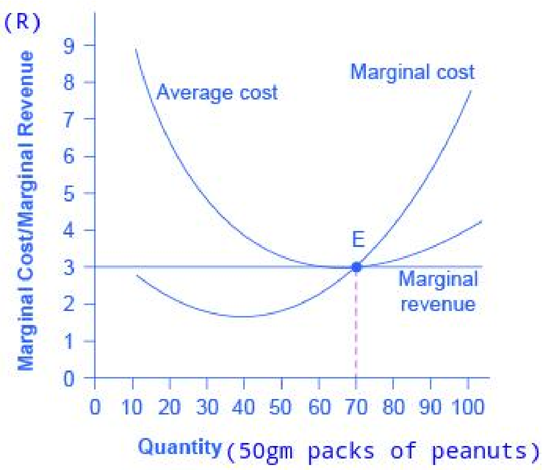
Why might a firm opt to continue operating in the immediate term (short term, the period in which there is at least one fixed cost) even if it is doing so at a loss? Well if the firm is running at a loss this means that some costs are not being covered by the incoming revenue. Costs can be split into fixed and variable costs. If times are bad (and everyone experiences bad times) maybe the payment of some costs can be delayed by negotiating with the suppliers of these inputs. Generally it is the fixed costs whose payments may be delayed. Thus, for example, loan/credit repayments could be rescheduled with banks or some suppliers. Rent and insurance payments could also be delayed by negotiation with the landlord and insurance companies. Variable costs, however, generally have to be paid on time (labour, most suppliers and utility providers such as electricity and water). One should also consider that it takes a lot of guts, sweat, tears and many years to get a successful business going. Why give up because of 2 or 3 months of losses? Things might always improve. This is why most businesses will generally try to push through hard times provided they can cover at least their variable costs until things improve.

Figure 5 illustrates three situations: (a) where price intersects marginal cost at a level above the average cost curve, (b) where price intersects marginal cost at a level equal to the average cost curve, and (c) where price intersects marginal cost at a level below the average cost curve.

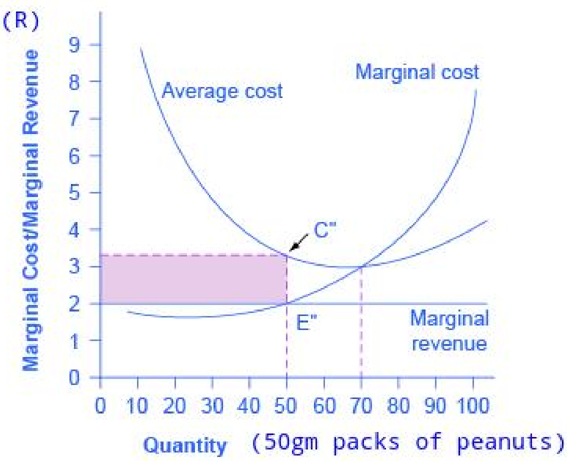
Price and Average Cost at The Peanut Farm



(a) Price is above average cost



(b) Price equals cost



(c) Price is below average cost

Figure 4: In (a), price intersects marginal cost above the average cost curve. Since price is greater than average cost, the firm is making a profit. In (b), price intersects marginal cost at the minimum point of the average cost curve. Since price is equal to average cost, the firm is breaking even. In (c), price intersects marginal cost below the average cost curve. Since price is less than average cost, the firm is making a loss.

First consider a situation where the price is equal to R5 for a pack of peanuts. The rule for a profit-maximizing perfectly competitive firm is to produce the level of output where Price= MR = MC, so the peanut farmer will produce a quantity of 90 packs as shown in Figure 5 (a). Remember that the area of a rectangle is equal to its base multiplied by its height. The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 90 packs

(the base) up to point E' (the height), over to the price of R5, and back to the origin. The average cost of producing 80 packs is shown by point C or about R3.50. Total costs will be the quantity of 80 times the average cost (cost of producing one pack) of R3.50, which is shown by the area of the rectangle from the origin to a quantity of 90, up to point C, over to the vertical axis and down to the origin. It should be clear from examining the two rectangles that total revenue is greater than total cost. Thus, profits will be the blue shaded rectangle on top.

It can be calculated as:
Equation:

$$\begin{aligned}\text{profit} &= \text{total revenue} - \text{total cost} \\ &= (90)(\text{R}5.00) - (90)(\text{R}3.50) \\ &= \text{R}135\end{aligned}$$

Or, it can be calculated as:
Equation:

$$\begin{aligned}\text{profit} &= (\text{price}\text{–}\,\text{average cost}) \times \text{quantity} \\ &= (\text{R}5.00\text{–}\,\text{R}3.50) \times 90 \\ &= \text{R}135\end{aligned}$$

Now consider Figure 5 (b), where the price has fallen to R3.00 for a 50gm pack of peanuts. Again, the perfectly competitive firm will choose the level of output where Price = MR = MC, but in this case, the quantity produced will be 70 packs. At this price and output level, where the marginal revenue, marginal cost and price lines intersect, the price received by the firm is exactly equal to its average cost of production (the MR, MC and ATC all intersect or meet at the same price and output level).

The farm’s total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 70 packs (the base) up to point E (the height), over to the price of R3, and back to the origin. The average cost of producing 70 packs is shown by point C’. Total costs will be the quantity of 70 packs times the average cost of R3.00, which is shown by the area of the rectangle from the origin to a quantity of 70 packs, up to point E, over to the vertical axis and down to the origin. It should be clear that the rectangles for total revenue and total cost are the

same. Thus, the firm is making zero economic profit. The calculations are as follows:
Equation:

$$\begin{aligned} \text{profit} &= \text{total revenue} - \text{total cost} \\ &= (70)(\text{R}3.00) - (70)(\text{R}3.00) \\ &= \text{R}0 \end{aligned}$$

Or, it can be calculated as:
Equation:

$$\begin{aligned} \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\ &= (\text{R}3.00 - \text{R}.00) \times 70 \\ &= \text{R}0 \end{aligned}$$

In Figure 5 (c), the market price has fallen still further to R2.00 for a pack of 50gm peanuts. At this price, marginal revenue intersects marginal cost at a quantity of 50 packs. The farm's total revenue at this price will be shown by the large shaded rectangle from the origin over to a quantity of 50 packs (the base) up to point E" (the height), over to the price of R2, and back to the origin. The average cost of producing 50 packs is shown by point C" or about R3.30. Total costs will be the quantity of 50 packs times the average cost of R3.30 per pack, which is shown by the area of the rectangle from the origin to a quantity of 50 packs, up to point C", over to the vertical axis and down to the origin. It should be clear from examining the two rectangles that total revenue is less than total cost. Thus, the firm is losing money and the loss (or negative economic profit) will be the rose-shaded rectangle.

The calculations are:
Equation:

$$\begin{aligned} \text{profit} &= (\text{total revenue} - \text{ total cost}) \\ &= (50)(\text{R}2.00) - (50)(\text{R}3.30) \\ &= -\text{R}77.50 \end{aligned}$$

Or:
Equation:

$$\begin{aligned} \text{profit} &= (\text{price} - \text{average cost}) \times \text{quantity} \\ &= (\text{R}1.75 - \text{R}3.30) \times 50 \\ &= -\text{R}77.50 \end{aligned}$$

If the market price received by a perfectly competitive firm leads it to produce at a quantity where the price is greater than average cost, the firm will earn profits. If the price received by the firm causes it to produce at a quantity where price equals average cost, which occurs at the minimum point of the AC curve, then the firm earns zero profits. Finally, if the price received by the firm leads it to produce at a quantity where the price is less than average cost, the firm will earn losses. This is summarized in Table 4.

Table 4: If...Then...

If...	Then...
Price > ATC	Firm earns an economic profit
Price = ATC	Firm earns zero economic profit
Price < ATC	Firm earns a loss

The Shutdown Point

The possibility that a firm may earn losses raises a question: Why can the firm not avoid losses by shutting down and not producing at all? The answer is that shutting down can reduce variable costs to zero, but in the short run, the firm has already paid for fixed costs. As a result, if the firm produces a quantity of zero, it would still make losses because it would still need to pay for its fixed costs. So, when a firm is experiencing losses, it must face a question: should it continue producing or should it shut down? We did discuss the reasons a firm may, under certain conditions, decide to continue operating even if it is making losses. But now let's take a real life example of when and how this might happen.

Vuyo's Gents Hair Salon will do as our example. Let's say that Vuyo signed a contract to rent space for his salon that costs R10,000 per month. If the firm decides to operate, its variable costs for hiring hairdressers is R15,000 for the month. If the firm shuts down, it must still pay the rent, but it would not need to hire labor. Three possible scenarios are shown below. In the first scenario, the salon does not have any clients, and therefore does not make any revenues, in

which case it faces losses of R10,000 equal to the fixed costs. In the second scenario, the salon has clients that earn Vuyo revenues of R10,000 for the month, but ultimately experiences losses of R15,000. In the third scenario, the salon earns revenues of R20,000 for the month, but experiences losses of R5,000.

In all three cases, Vuyo's Salon loses money. In all three cases, when the rental contract expires in the long run, assuming revenues do not improve, the firm should exit this business. In the short run, though, the decision varies depending on the level of losses and whether the firm can cover its variable costs (as discussed earlier in the chapter). In scenario 1, the salon does not have any revenues, so hiring hairdressers would increase variable costs and losses, so it should shut down and only incur its fixed costs. In scenario 2, the salon's losses are greater because it does not make enough revenue to offset the increased variable costs plus fixed costs, so it should shut down immediately. If price is below the minimum average variable cost, the firm must shut down. This is the **shut-down rule** that firms will generally try to follow. In contrast, in scenario 3 the revenue that the salon can earn is high enough that the losses diminish when it remains open, so the salon should remain open in the short run.

Scenario 1
If the Salon shuts down now, revenues are zero but it will not incur any variable costs and would only need to pay fixed costs of R10,000
Equation:

$$\begin{aligned}&\text{profit} = \text{total revenue} – (\text{fixed costs} + \text{variable cost})\\&= 0 – \text{R}10{,}000\\&= –\text{R}10{,}000\end{aligned}$$

Scenario 2
The Salon earns revenues of R10,000, and variable costs are R15,000. The salon should shut down now.
Equation:

$$\begin{aligned}&\text{profit} = \text{total revenue} – (\text{fixed costs} + \text{variable cost})\\&= \text{R}10{,}000 – (\text{R}10{,}000 + \text{R}15{,}000)\\&= –\text{R}15{,}000\end{aligned}$$

Scenario 3

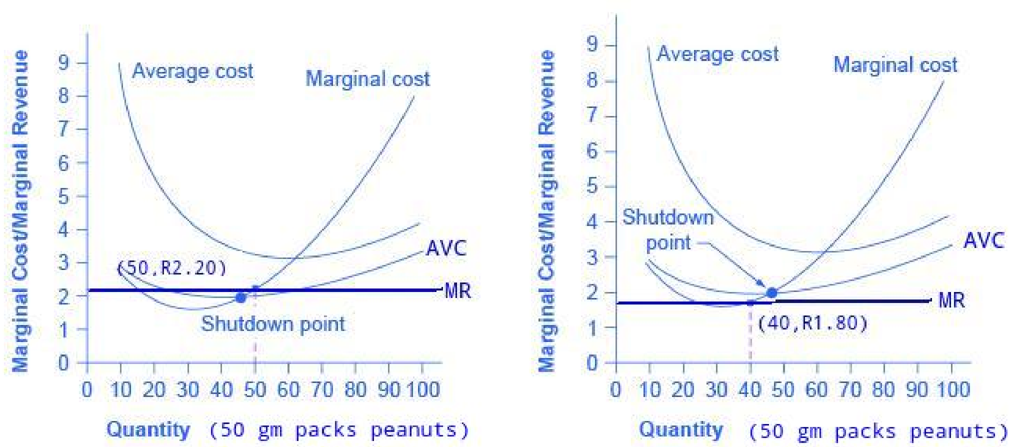
The salon earns revenues of R20,000, and variable costs are R15,000. The salon should continue in business.

Equation:

$$\begin{aligned} \text{profit} &= \text{total revenue} – (\text{fixed costs} + \text{variable cost}) \\ &= \text{R}20{,}000 – (\text{R}10{,}000 + \text{R}15{,}000) \\ &= –\text{R}5{,}000 \end{aligned}$$

This example suggests that the key factor is whether a firm can earn enough revenues to cover at least its variable costs by remaining open. Let’s return now to our small-scale peanut farm. Figure 5 illustrates this lesson by adding the average variable cost curve to the marginal cost and average cost curves. At a price of R2.20 per pack, as shown in Figure 6 (a), the farm produces at a level of 50 packs. It is making losses of R56 (as explained earlier), but price is above average variable cost and so the farm continues to operate. However, if the price declined to R1.80 per pack, as shown in Figure 6 (b), and if the farm applied its rule of producing where P = MR = MC, it would produce a quantity of 40 packs. This price is below average variable cost for this level of output. If the farmer cannot pay workers (the variable costs), then it has to shut down. At this price and output, total revenues would be R72 (quantity of 40 times price of R1.80) and total cost would be R144, for overall losses of R72. If the farm shuts down, it must pay only its fixed costs of R62, so shutting down is preferable to selling at a price of R1.80 per pack.

The Shutdown Point for the Peanut Farm



(a) Price is above average variable cost (b) Price is below average variable cost

Figure 6: In (a), the farm produces at a level of 50 packs. It is making losses of R56, but price is above average variable cost, so it continues to operate. In (b), total revenues are R72 and total cost is R144, for overall losses of R72. If the farm shuts down, it must pay only its fixed costs of R62. Shutting down is preferable to selling at a price of R1.80 per pack.

Looking at Table 5 below, if the price falls below R2.05, the minimum average variable cost, the farm must shut down.

Table 5: Cost of Production for the Peanut Farm

Quantity (50gm packs)	Total Cost [R]	Fixed Cost [R]	Variable Cost [R]	Marginal Cost [R]	Average Cost [R]	Average Variable Cost [R]
0	62	62	-	-	-	-
				2.8		
10	90	62	28		9	2.8
20	110	62	48		5.5	2.4
				1.6		
30	126	62	64		4.2	2.13
				1.8		
40	144	62	82		3.6	2.05
				2.2		
50	166	62	104		3.32	2.08
				2.6		
60	192	62	130		3.2	2.16
				3.2		
70	224	62	162		3.2	2.31
				4		
80	264	62	202		3.3	2.52
				6		
90	324	62	262		3.6	2.91
				8		
100	404	62	342		4.04	3.42

The intersection of the average variable cost, marginal cost and marginal revenue curves, which shows the price where the firm would lack enough revenue to cover its variable costs, is called the **shutdown point**. If the perfectly competitive firm can charge a price above the shutdown point, then the firm is at least covering its average variable costs. In other words the price received for the product at least covers the variable costs of producing that item. It is also making enough revenue

to cover at least a portion of fixed costs, so it should limp ahead even if it is making losses in the short run, since at least those losses will be smaller than if the firm shuts down immediately and incurs a loss equal to total fixed costs. However, if the firm is receiving a price below the price at the shutdown point, then the firm is not even covering its variable costs. In this case, staying open is making the firm's losses larger, and it should shut down immediately. To summarize, if:

- price < minimum average variable cost, then firm shuts down
- price = minimum average variable cost, then firm stays in business

Short-Run Outcomes for Perfectly Competitive Firms

The average cost and average variable cost curves divide the marginal cost curve into three segments, as shown in Figure 7. At the market price, which the perfectly competitive firm accepts as given, the profit-maximizing firm chooses the output level where price or marginal revenue, which are the same thing for a perfectly competitive firm, is equal to marginal cost: P = MR = MC.

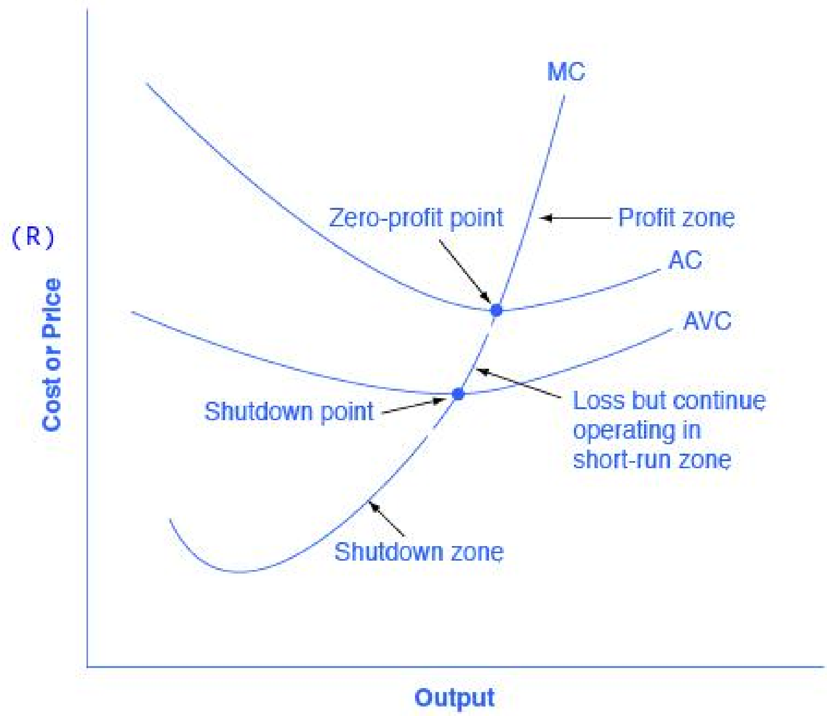


Figure 7: The marginal cost curve can be divided into

three zones, based on where it is crossed by the average cost and average variable cost curves. The point where MC crosses AC is called the zero-profit point. If the firm is operating at a level of output where the market price is at a level higher than the zero-profit point, then price will be greater than average cost and the firm is earning profits. If the price is exactly at the zero-profit point, then the firm is making zero profits. If price falls in the zone between the shutdown point and the zero-profit point, then the firm is making losses but will continue to operate in the short run, since it is covering its variable costs. However, if price falls below the price at the shutdown point, then the firm will shut down immediately, since it is not even covering its variable costs.

First consider the upper zone, where prices are above the level where marginal cost (MC) crosses average cost (AC) at the zero profit point. At any price above that level, the firm will earn profits in the short run. If the price falls exactly on the zero profit point where the MC and AC curves cross, then the firm earns zero profits. If a price falls into the zone between the zero profit point, where MC crosses AC, and the shutdown point, where MC crosses AVC, the firm will be making losses in the short run—but since the firm is more than covering its variable costs, the losses are smaller than if the firm shut down immediately. Finally, consider a price at or below the shutdown point where MC crosses AVC. At any price like this one, the firm will shut down immediately, because it cannot even cover its variable costs.

Marginal Cost and the Firm's Supply Curve

For a perfectly competitive firm, the marginal cost curve is identical to the firm's supply curve starting from the minimum point on the average variable cost curve. To understand why this is, first think about what the supply curve means. A firm checks the market price and then looks at its supply curve to decide what quantity to produce. Now, think about what it means to say that a firm will maximize its profits by producing at the quantity where P = MC. This rule means that the firm checks the market price, and then looks at its marginal cost to determine the quantity to produce—and makes sure that the price is greater than the minimum

average variable cost. In other words, the marginal cost curve above the minimum point on the average variable cost curve becomes the firm's supply curve.

As discussed in the chapter on Demand and Supply many of the reasons that supply curves shift relate to underlying changes in costs. For example, a lower price of key inputs or new technologies that reduce production costs cause supply to shift to the right; in contrast, bad weather or added government regulations can add to costs of certain goods in a way that causes supply to shift to the left. These shifts in the firm's supply curve can also be interpreted as shifts of the marginal cost curve. A shift in costs of production that increases marginal costs at all levels of output—and shifts MC to the left—will cause a perfectly competitive firm to produce less at any given market price. Conversely, a shift in costs of production that decreases marginal costs at all levels of output will shift MC to the right and as a result, a competitive firm will choose to expand its level of output at any given price. The following Work It Out feature will walk you through an example.

Note:

At What Price Should the Firm Continue Producing in the Short Run?

To determine the short-run economic condition of a firm in perfect competition, follow the steps outlined below. Use the data shown in Table 6 below. All costs, revenues and profits are in Rands.

Table 6: Costs, Revenues and Profits

Q	P	TFC	TVC	TC	AVC	ATC	MC	TR	Profits
0	28	20	0	-	-	-	-	-	-
1	28	20	20	-	-	-	-	-	-
2	28	20	25	-	-	-	-	-	-
3	28	20	35	-	-	-	-	-	-
4	28	20	52	-	-	-	-	-	-
5	28	20	80	-	-	-	-	-	-

Step 1. Determine the cost structure for the firm. For a given total fixed costs and variable costs, calculate total cost, average variable cost, average total cost, and

marginal cost. Follow the formulas given in the Cost and Industry Structure chapter. These calculations are shown in Table 7 below. All costs are in Rands.

Table 7: Output Levels, Prices and Costs

Q	P	TFC	TVC	TC (TFC+TVC)	AVC (TVC/Q)	ATC (TC/Q)	MC (TC2−TC1)/(Q2−Q1)
0	28	20	0	20+0=20	-	-	-
							(40−20)/(1−0)= 20
1	28	20	20	20+20=40	20/1=20.00	40/1=40.00	
							(45−40)/(2−1)= 5
2	28	20	25	20+25=45	25/2=12.50	45/2=22.50	
							(55−45)/(3−2)= 10
3	28	20	35	20+35=55	35/3=11.67	55/3=18.33	
							(72−55)/(4−3)= 17
4	28	20	52	20+52=72	52/4=13.00	72/4=18.00	
							(100−72)/(5−4)= 28
5	28	20	80	20+80=100	80/5=16.00	100/5=20.00	

Step 2. Determine the market price that the firm receives for its product. This should be given information, as the firm in perfect competition is a price taker. With the given price, calculate total revenue as equal to price multiplied by quantity for all output levels produced. In this example, the given price is R28. You can see that in the second column of Table 8.

Step 3. Calculate profits as total cost subtracted from total revenue, as shown in Table 9.

Table 9: Total Costs, Total Revenues and Total Profits

Quantity	Total Revenue [R]	Total Cost [R]	Profits (TR−TC) [R]
0	0	20	0−20=−20
1	28	40	28−40=−12
2	56	45	56−45=11
3	84	55	84−55=29
4	112	72	112−72=40
5	140	100	140−100=40

Step 4. To find the profit-maximizing output level, look at the Marginal Cost column (at every output level produced), as shown in Table 10, and determine where it is equal to the market price. The output level where price equals the marginal cost is the output level that maximizes profits.

Table 10: Costs, Revenues and Profits

Q	P [R]	TFC[R]	TVC[R]	TC[R]	AVC[R]	ATC[R]	MC[R]	TR[R]	Profits[R]
0	28	20	0	20	-	-	-	0	−20
1	28	20	20	40	20	40	20	28	−12
2	28	20	25	45	12.5	22.5	5	56	11
3	28	20	35	55	11.67	18.33	10	84	29
4	28	20	52	72	13	18	17	112	40
5	**28**	20	80	100	16.4	20.4	**30**	140	**40**

Step 5. Once you have determined the profit-maximizing output level (in this case, output quantity 5), you can look at the amount of profits made (in this case, R40).

Step 6. If the firm is making economic losses, the firm needs to determine whether it produces the output level where price equals marginal revenue and equals marginal cost or it shuts down and only incurs its fixed costs.

Step 7. For the output level where marginal revenue is equal to marginal cost, check if the market price is greater than the average variable cost of producing that output level.

- If P > AVC but P < ATC, then the firm continues to produce in the short-run, making economic losses.
- If P < AVC, then the firm stops producing and only incurs its fixed costs.

In this example, the price of R28 is greater than the AVC (R16.40) as well as the ATC (R20.40) of producing 5 units of output resulting in economic profits, so in

this case the firm will certainly continue to operate.

Key Concepts and Summary

As a perfectly competitive firm produces a greater quantity of output, its total revenue steadily increases at a constant rate determined by the given market price. Profits will be highest (or losses will be smallest) at the quantity of output where total revenues exceed total costs by the greatest amount (or where total revenues fall short of total costs by the smallest amount). Alternatively, profits will be highest where marginal revenue, which is price for a perfectly competitive firm, is equal to marginal cost. If the market price faced by a perfectly competitive firm is above average cost at the profit-maximizing quantity of output, then the firm is making profits. If the market price is below average cost at the profit-maximizing quantity of output, then the firm is making losses.

If the market price is equal to average cost at the profit-maximizing level of output, then the firm is making zero economic profits (however it will be earning normal profits). The point where the marginal cost and marginal revenue curves cross the average cost curve, at the minimum of the average cost curve, is called the “zero economic profit point.” If the market price faced by a perfectly competitive firm is below average variable cost at the profit-maximizing quantity of output, then the firm should shut down operations immediately. If the market price faced by a perfectly competitive firm is above average variable cost, but below average cost, then the firm should continue producing in the short run, but exit in the long run since firms need to ultimately cover ALL costs and not just a portion of the costs. The point where the marginal cost and marginal revenue curves cross the average variable cost curve below its minimum point is called the shutdown point.

Self-Check Questions

Exercise 1

What would happen to our small-scale peanut farmer's profits if the market price for a 50gm pack of peanuts increases from R4 per pack to R6 per pack of peanuts?

Table 11: Costs, Revenues and Profits at R6/pack

Quantity (50gm packs)	Total Cost [R]	Fixed Cost [R]	Variable Cost [R]	Total Revenue [R]	Profit [R]
0	62	62	-	0	−62
10	90	62	28	60	−30
20	110	62	48	120	10
30	126	62	64	180	54
40	144	62	82	240	96
50	166	62	104	300	134
60	192	62	130	360	168
70	224	62	162	420	196
80	264	62	202	480	216
90	324	62	262	540	216
100	404	62	342	600	196

Solution
Holding costs constant, profits would increase as shown in Table 11

Exercise 2
Suppose that the market price increases to R6, as shown in Table 12. What would happen to the profit-maximizing output level?

Table 12: Costs, Revenues and Profits at R6/pack

Quantity (50gm packs)	**Total Cost[R]**	**Fixed Cost[R]**	**Variable Cost[R]**	**Marginal Cost[R]**	**Total Revenue[R]**	**Marginal Revenue[R]**
0	62	62	-	-	0	-
				2.8		6
10	90	62	28		60	
				2		6
20	110	62	48		120	
				1.6		6
30	126	62	64		180	
				1.8		6
40	144	62	82		240	
				2.2		6
50	166	62	104		300	
				2.6		6
60	192	62	130		360	
				3.2		6
70	224	62	162		420	
				4		6
80	264	62	202		480	
				6		**6**
90	324	62	262		540	
				8		6
100	404	62	342		600	

Solution
Profit maximising output level would increase to between 80 - 90 packs peanuts i.e. where MC = MR.

Exercise 3
Explain in words why a profit-maximizing firm will not choose to produce at a quantity where marginal cost exceeds marginal revenue.

Solution
If marginal costs exceeds marginal revenue, then the firm will reduce its profits for every additional unit of output it produces. Profit would be greatest if it reduces output to where MR = MC.

Exercise 4
A firm’s marginal cost curve above the average variable cost curve is equal to the firm’s individual supply curve. This means that every time a firm receives a price from the market it will be willing to supply the amount of output where the price

equals marginal cost. What happens to the firm’s individual supply curve if marginal costs increase?

Solution
The firm will be willing to supply fewer units at every price level. In other words, the firm’s individual supply curve decreases and shifts to the left.

Review Questions

Exercise 5
How does a perfectly competitive firm decide what price to charge?

Exercise 6
What prevents a perfectly competitive firm from seeking higher profits by increasing the price that it charges?

Exercise 7
How does a perfectly competitive firm calculate total revenue?

Exercise 8
Briefly explain the reason for the shape of a marginal revenue curve for a perfectly competitive firm.

Exercise 9
What two rules does a perfectly competitive firm apply to determine its profit-maximizing quantity of output?

Exercise 10
How does the average cost curve help to show whether a firm is making profits or losses?

Exercise 11
What two lines on a cost curve diagram intersect at the zero-profit point?

Exercise 12
Should a firm shut down immediately if it is making losses?

Exercise 13
How does the average variable cost curve help a firm know whether it should shut down immediately?

Exercise 14
What two lines on a cost curve diagram intersect at the shutdown point?

Critical Thinking Questions

Exercise 15
Your company operates in a perfectly competitive market. You have been told that advertising can help you increase your sales in the short run. Would you create an aggressive advertising campaign for your product?

Exercise 16
Since a perfectly competitive firm can sell as much as it wishes at the market price, why can the firm not simply increase its profits by selling an extremely high quantity?

Problems

Exercise 17
Samkelo sells bags he makes at home to university students to carry their files. His fixed costs of production are R20 (a small rental he pays his mum who loves him too much to charge more). The total variable costs are R20 for one bag, R25 for two units, R35 for the three units, R50 for four units, and R80 for five units. In the form of a table, calculate total revenue, marginal revenue, total cost, and marginal cost for each output level (one to five units). What is the profit-maximizing quantity of output? On one diagram, sketch the total revenue and total cost curves. On another diagram, sketch the marginal revenue and marginal cost curves.

Exercise 18
Zeek Boutique, a new outlet at Liberty Midlands Mall in Pietermaritzburg, sells a new range of winter jackets for ladies. Jackets sell for R720 each. The fixed costs of production are R1000 per month (a low rental in the first year to attract new tenants). The total variable costs are R640 for one unit, R840 for two units, R1 140 for three units, R1 840 for four units, and R2 700 for five units. In the form of a table, calculate total revenue, marginal revenue, total cost and marginal cost for each output level (one to five units). On one diagram, sketch the total revenue and total cost curves. On another diagram, sketch the marginal revenue and marginal cost curves. What is the profit maximizing quantity?

Exercise 19

A small home-based Pinetown company produces affordable, easy-to-use home computer systems and has fixed costs of R2 500. The marginal cost of producing computers is R7 000 for the first computer, R2 500 for the second, R3 000 for the third, R3 500 for the fourth, R4 000 for the fifth, R4 500 for the sixth, and R5 000 for the seventh.

a. Create a table that shows the company's output, total cost, marginal cost, average cost, variable cost, and average variable cost.
b. At what price is the zero-profit point? At what price is the shutdown point?
c. If the company sells the computers for R5 000, is it making a profit or a loss? How big is the profit or loss? Sketch a graph with AC, MC, and AVC curves to illustrate your answer and show the profit or loss.
d. If the firm sells the computers for R3 000, is it making a profit or a loss? How big is the profit or loss? Sketch a graph with AC, MC, and AVC curves to illustrate your answer and show the profit or loss.

Glossary

marginal revenue
: the additional revenue gained from selling one more unit

shutdown point
: level of output where the marginal cost curve intersects the average variable cost curve at the minimum point of AVC; if the price is below this point, the firm should shut down immediately

Entry and Exit Decisions in the Long Run

By the end of this section, you will be able to:

- Explain how entry and exit lead to zero profits in the long run
- Discuss the long-run adjustment process

The line between the short run and the long run cannot be defined precisely with a stopwatch, or even with a calendar. It varies according to the specific business. The distinction between the short run and the long run is therefore more technical: in the short run, firms cannot change the usage of fixed inputs, while in the long run, the firm can adjust all factors of production.

In a competitive market, profits serve to attract other firms (competition!). If a business is making a profit in the short run, it has an incentive to expand existing factories or to build new ones. New firms may start production, as well. When new firms enter the industry in response to increased industry profits it is called **entry**.

Losses are the black thundercloud that causes businesses to flee. If a business is making losses in the short run, it will either keep limping along or just shut down, depending on whether its revenues are covering its variable costs. But in the long run, firms that are facing losses will shut down at least some of their output, and some firms will cease production altogether. The long-run process of reducing production in response to a sustained pattern of losses is called **exit**. The following Clear It Up feature discusses where some of these losses might come from, and the reasons why some firms go out of business.

Note:

Why do firms cease to exist?

Can we say anything about what causes a firm to exit an industry? Profits are the measurement that determines whether a business stays operating or not. Individuals start businesses with the purpose of making profits. They invest their money, time, effort, and many other resources to produce and sell something that they hope will give them something in return.

Unfortunately, not all businesses are successful, and many new startups soon realize that their “business adventure” must eventually end.

In the model of perfectly competitive firms, those that consistently cannot make money will “exit,” which is a nice word for a more painful process. When a business fails, after all, workers lose their jobs, investors lose their money, and owners and managers can lose their dreams. Many businesses fail. Between 2008 and 2012, 440 000 small businesses in South Africa closed their doors primarily as a consequence of the recession which commenced in 2009 (It’s Been a Bad, Bad Time for Small Businesses: 2012).

Sometimes a business fails because of poor management, or workers who are not very productive, or because of tough domestic or foreign competition. Standard Bank (2014) notes that 50% of South African small businesses fail within the first two years due to lack of ability, inexperience, poor planning, inability to access finance and failure to exercise adequate financial controls. Businesses also fail from a variety of causes that might best be summarized as bad luck. For example, conditions of demand and supply in the market shift in an unexpected way, so that the prices that can be charged for outputs fall, or the prices that need to be paid for inputs rise. Small businesses in South Africa account for 50% of all employment and about 45% of the country's Gross Domestic Product or GDP (Industrial Development Corporation: 2014). So even a small fraction of them failing will affect many people — and business failures can be very hard on the workers and managers directly involved. But from the standpoint of the overall economic system, business exits are sometimes a necessary evil if a market-oriented system is going to offer a flexible mechanism for satisfying customers, keeping costs low, and inventing new products.

How Entry and Exit Lead to Zero Profits in the Long Run

No perfectly competitive firm acting alone can affect the market price. However, the combination of many firms entering or exiting the market will affect overall supply in the market. In turn, a shift in supply for the market as a whole will affect the market price. Entry and exit to and from the

market are the driving forces behind a process that, in the long run, pushes the price down to minimum average total costs so that all firms are earning a zero economic profit (but at least earning normal profit).

To understand how short-run profits for a perfectly competitive firm will disappear in the long run, imagine the following situation. The market is in **long-run equilibrium**, where all firms earn zero economic profits producing the output level where P = MR = MC and P = AC. No firm has the incentive to enter or leave the market. Let's say that the product's demand increases, and with that, the market price goes up. The existing firms in the industry are now facing a higher price than before, so they will increase production to the new output level where P = MR = MC.

This will temporarily make the market price rise above the average cost curve, and therefore, the existing firms in the market will now be earning economic profits. However, these economic profits attract other firms to enter the market. Entry of many new firms causes the market supply curve to shift to the right. As the supply curve shifts to the right, the market price starts decreasing, and with that, economic profits fall for new and existing firms. As long as there are still economic profits in the market, entry will continue to shift supply to the right. This will stop whenever the market price is driven down to the zero-economic profit level, where no firm is earning economic profits (but earning only normal profits).

Short-run losses will fade away by reversing this process. Say that the market is in long-run equilibrium (firms are earning only normal profits). This time, instead, market demand decreases, and with that, the market price starts falling. The existing firms in the industry are now facing a lower price than before, and as it will be below the average cost curve, they will now be making economic losses. Some firms will continue producing where the new P = MR = MC, as long as they are able to cover their average variable costs. Some firms will have to shut down immediately as they will not be able to cover their average variable costs, and will then only incur their fixed costs, minimizing their losses. Exit of many firms causes the market supply curve to shift to the left (decrease). As the supply curve shifts to the left, the market price starts rising, and economic losses start to be lower. This process ends whenever the market price rises to the

zero-economic profit level, where the existing firms are no longer losing money and are at zero economic profits again. Firms that survive are prepared to remain in operation even if their economic profits are zero since they will at least be earning normal profits.

Thus, while a perfectly competitive firm can earn profits in the short run, in the long run the process of entry will push down prices until they reach the zero-profit level. Conversely, while a perfectly competitive firm may earn losses in the short run, firms will not continually lose money. In the long run, firms making losses are able to escape from their fixed costs and their exit from the market will push the price back up to the zero-profit level. In the long run, this process of entry and exit (remember that a condition of perfect competition is ease of entry to or exit from an industry) will drive the price in perfectly competitive markets to the zero-economic profit point at the bottom of the AC curve, where marginal cost crosses average cost.

The Long-Run Adjustment and Industry Types

Whenever there are expansions in an industry, costs of production for the existing and new firms could either stay the same, increase, or even decrease. Therefore, we can categorize an industry as being (1) a constant cost industry (as demand increases, the cost of production for firms stays the same), (2) an increasing cost industry (as demand increases, the cost of production for firms increases), or (3) a decreasing cost industry (as demand increases the costs of production for the firms decreases).

For a constant cost industry, whenever there is an increase in market demand and price, then the supply curve shifts to the right with new firms' entry and stops at the point where the new long-run equilibrium intersects at the same market price as before. But why will costs remain the same? In this type of industry, the supply curve is very elastic. Firms can easily supply any quantity that consumers demand. In addition, there is a perfectly elastic supply of inputs—firms can easily increase their demand for employees, for example, with no increase to wages. An increased demand for ethanol (used to make synthetic fuel) in recent years has caused the demand for maize to increase. Consequently, many farmers switched from

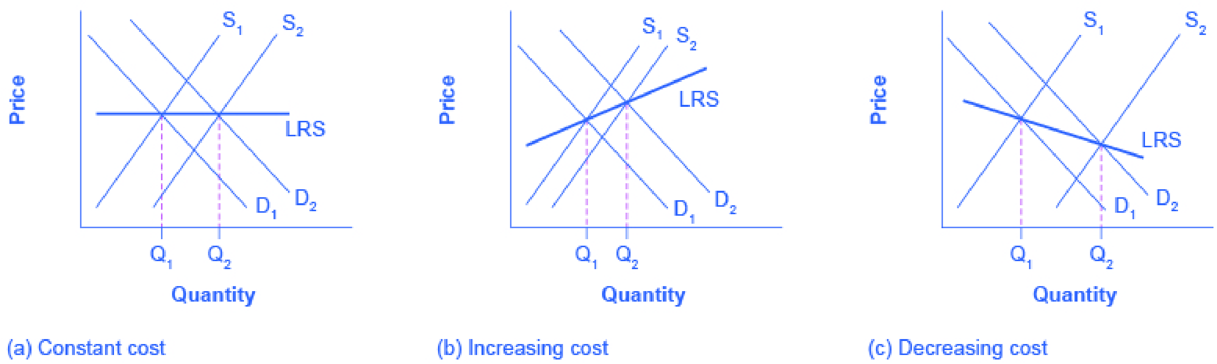
growing wheat to growing maize. Agricultural markets are generally good examples of constant cost industries.

For an increasing cost industry, as the market expands, the old and new firms experience increases in their costs of production, which makes the new zero-profit level intersect at a higher price than before. Here companies may have to deal with limited inputs, such as skilled labor. As the demand for these workers rise, wages rise and this increases the cost of production for all firms. The industry supply curve in this type of industry is more inelastic. Most industries would probably fall into this category, for example motor vehicle manufacturing.

For a decreasing cost industry, as the market expands, the old and new firms experience lower costs of production, which makes the new zero-profit level intersect at a lower price than before. In this case, the industry and all the firms in it are experiencing falling average total costs. This can be due to an improvement in technology in the entire industry or an increase in the education of employees. High tech industries (for example computer components such as memory chips) may be a good example of a decreasing cost market.

Figure 1 (a) presents the case of an adjustment process in a constant cost industry. Whenever there are output expansions in this type of industry, the long-run outcome implies more output produced at exactly the same original price. Note that supply was able to increase to meet the increased demand. When we join the before and after long-run equilibriums, the resulting line is the long run supply (LRS) curve in perfectly competitive markets. In this case, it is a flat curve. Figure 1 (b) and Figure 1 (c) present the cases for an increasing cost and decreasing cost industry, respectively. For an increasing cost industry, the LRS is upward sloping, while for a decreasing cost industry, the LRS is downward sloping.

Adjustment Process in different types of Industries



In (a), demand increased and supply met it. Notice that the supply increase is equal to the demand increase. The result is that the equilibrium price stays the same as quantity sold increases. In (b), notice that sellers were not able to increase supply as much as demand. Some inputs were scarce, or wages were rising. The equilibrium price rises. In (c), sellers easily increased supply in response to the demand increase. Here, new technology or economies of scale caused the large increase in supply, resulting in declining equilibrium price.

Key Concepts and Summary

In the long run, firms will respond to profits through a process of entry (ease of entry is one of the conditions of perfect competition), where existing firms expand output and new firms enter the market. Conversely, firms will react to losses in the long run through a process of exit (ease of exit is one of the assumptions of perfect competition), in which existing firms reduce output or cease production altogether. Through the process of entry in response to economic profits and exit in response to losses, the price level in a perfectly competitive market will move toward the zero-economic profit point, where the marginal cost curve crosses the AC curve, at the minimum of the average cost curve.

The long-run supply curve shows the long-run output supplied by firms in three different types of industries: constant cost, increasing cost, and decreasing cost.

Self-Check Questions

Exercise:

Problem:

If new technology in a perfectly competitive market brings about a substantial reduction in costs of production, how will this affect the market?

Solution:

With a technological improvement that brings about a reduction in costs of production, an adjustment process will take place in the market. The technological improvement will result in an increase in supply curves, by individual firms and at the market level. The existing firms will experience higher profits for a while, which will attract other firms into the market. This entry process will stop whenever the market supply increases enough (both by existing and new firms) so profits are driven back to zero.

Exercise:

Problem:

A market in perfect competition is in long-run equilibrium (earning only normal profits). What happens to the market if labor unions are able to increase wages for workers?

Solution:

When wages increase, costs of production increase. Some firms would now be making economic losses and would shut down. The supply curve then starts shifting to the left, pushing the market price up. This process ends when all firms remaining in the market earn zero economic profits. The result is a contraction in the output produced in the market.

Review Questions

Exercise:

Problem: Why does entry occur?

Exercise:

Problem: Why does exit occur?

Exercise:

Problem:

Do entry and exit occur in the short run, the long run, both, or neither?

Exercise:

Problem:

What price will a perfectly competitive firm end up charging in the long run? Why?

Critical Thinking Questions

Exercise:

Problem:

Many firms in South Africa fail to make their profit targets each month and many more run at a loss, yet they still continue operating. Why would they do this instead of completely shutting down?

Exercise:

Problem:

Why will profits for firms in a perfectly competitive industry tend to vanish in the long run?

Exercise:

Problem:

Why will losses for firms in a perfectly competitive industry tend to vanish in the long run?

References

Industrial Development Corporation. 2014. Small business, huge potential for South Africa. IDC Media Room, October 2. Available http://www.idc.co.za/home/media-room/articles/738-small-business,-huge-potential-for-south-africa.html (Accessed: 2 May 2016)

It's been a bad, bad time for small businesses. 2012. Mail and Guardian Online, Business, March 12. Available http://mg.co.za/article/2012-03-12-employment-index-shows-record-low-in-business-startups (Accessed: 2 May 2016)

Standard Bank. 2014. Common reasons why businesses fail in South Africa. BizConnect, October 9. Available https://bizconnect.standardbank.co.za/start/business-planning/common-reasons-why-businesses-fail-in-south-africa.aspx (Accessed: 2 May 2016)

Glossary

entry
: the long-run process of firms entering an industry in response to industry profits

exit
: the long-run process of firms reducing production and shutting down in response to industry losses

long-run equilibrium
: where all firms earn zero economic profits producing the output level where P = MR = MC and P = AC

Efficiency in Perfectly Competitive Markets

By the end of this section, you will be able to:

- Apply concepts of productive efficiency and allocative efficiency to perfectly competitive markets
- Compare the model of perfect competition to real-world markets

When profit-maximizing firms in perfectly competitive markets combine with utility-maximizing consumers, something remarkable happens: the resulting quantities of outputs of goods and services demonstrate both productive and allocative efficiency (terms that were first introduced in Choice in a World of Scarcity).

Productive efficiency means producing without waste, so that the choice is on the production possibility frontier. In the long run in a perfectly competitive market, because of the process of entry and exit, the price in the market is equal to the minimum of the long-run average cost curve. In other words, goods are being produced and sold at the lowest possible average cost.

Allocative efficiency means that among the points on the production possibility frontier, the point that is chosen is socially preferred—at least in a particular and specific sense. In a perfectly competitive market, price will be equal to the marginal cost of production. Think about the price that is paid for a good as a measure of the social benefit (value) received for that good; after all, willingness to pay conveys what the good is worth to a buyer. Then think about the marginal cost of producing the good as representing not just the cost for the firm, but more broadly as the social cost of producing that good. When perfectly competitive firms follow the rule that profits are maximized by producing at the quantity where price is equal to marginal cost, they are thus ensuring that the social benefits received from producing a good are in line with the social costs of production.

Figure 1

To explore what is meant by allocative efficiency, it is useful to walk through an example. Begin by assuming that the market for wholesale flowers is perfectly competitive, and so P = MC. Now, consider what it would mean if firms in that market produced a lesser quantity of flowers. At a lesser quantity, marginal costs will not yet have increased as much, so that price will exceed marginal cost; that is, P > MC. In that situation, the benefit to society as a whole of producing additional goods, as measured by the willingness of consumers to pay for marginal units of a good, would be higher than the cost of the inputs of labor and physical capital needed to produce the marginal good. In other words, the gains to society as a whole from producing additional marginal units will be greater than the costs.

Conversely, consider what it would mean if, compared to the level of output at the allocatively efficient choice when P = MC, firms produced a greater quantity of flowers. At a greater quantity, marginal costs of production will have increased so that P < MC. In that case, the marginal costs of producing additional flowers is greater than the benefit to society as measured by what people are willing to pay. For society as a whole, since the costs are outstripping the benefits, it will make sense to produce a lower quantity of such goods.

When perfectly competitive firms maximize their profits by producing the quantity where P = MC, they also assure that the benefits to consumers of what they are buying, as measured by the price they are willing to pay, is equal to the costs to society of producing the marginal units, as measured by the marginal costs the firm must pay—and thus that allocative efficiency holds.

The statement that a perfectly competitive market in the long run will feature both productive and allocative efficiency is not entirely true. Remember, economists are using the concept of “efficiency” in a particular and specific sense, not as a synonym for “desirable in every way.” For one thing, consumers’ ability to pay reflects the income distribution in a particular society. Thus, a homeless person may have no ability to pay for housing because they have insufficient income.

Perfect competition is only an ideal and is not realistic. If it was in fact attainable it would be of enormous benefit to society and consumers in particular. However, for market structures such as monopoly, monopolistic competition, and oligopoly, which are more frequently observed in the real world than perfect competition, firms will not always produce at the minimum of average cost, nor will they always set price equal to marginal cost. Thus, these other competitive situations will not produce productive and allocative efficiency.

Moreover, real-world markets include many serious issues that are discounted or ignored in the model of perfect competition, including pollution, inventions of new technology, poverty which may make some people unable to pay for basic necessities of life, government programs like national defence or education, discrimination in labor markets, and buyers

and sellers who must deal with imperfect and unclear information. These issues are explored in other chapters. However, the theoretical efficiency of perfect competition does provide a useful measure or benchmark for comparing the issues that arise from these real-world problems.

Key Concepts and Summary

Long-run equilibrium in perfectly competitive markets meets two important conditions: allocative efficiency and productive efficiency. These two conditions have important implications. First, resources are allocated to their best alternative use. Second, they provide the maximum satisfaction attainable by society.

Self-Check Questions

Exercise:

Problem:

Productive efficiency and allocative efficiency are two concepts achieved in the long run in a perfectly competitive market. These are the two reasons why we call them “perfect.” How would you use these two concepts to analyze other market structures and label them “imperfect?”

Solution:

Perfect competition is considered to be “perfect” because both allocative and productive efficiency are met at the same time in a long-run equilibrium. If a market structure results in long-run equilibrium that does not minimize average total costs and/or does not charge a price equal to marginal cost, then either allocative or productive (or both) efficiencies are not met, and therefore the market cannot be labeled “perfect.”

Exercise:

Problem:

Explain how the profit-maximizing rule of setting P = MC leads a perfectly competitive market to be allocatively efficient.

Solution:

Think of the market price as representing the gain to society from a purchase, since it represents what someone is willing to pay. Think of the marginal cost as representing the cost to society from making the last unit of a good. If P > MC, then the benefits from producing more of a good exceed the costs, and society would gain from producing more of the good. If P < MC, then the social costs of producing the marginal good exceed the social benefits, and society should produce less of the good. Only if P = MC, the rule applied by a profit-maximizing perfectly competitive firm, will society's costs and benefits be in balance. This choice will be the option that brings the greatest overall benefit to society.

Review Questions

Exercise:

Problem:

Will a perfectly competitive market display productive efficiency? Why or why not?

Exercise:

Problem:

Will a perfectly competitive market display allocative efficiency? Why or why not?

Critical Thinking Questions

Exercise:

Problem:

Assuming that the market for cigarettes is in perfect competition, what does allocative and productive efficiency imply in this case? What does it not imply?

Exercise:

Problem:

In the argument for why perfect competition is allocatively efficient, the price that people are willing to pay represents the gains to society and the marginal cost to the firm represents the costs to society. Can you think of some social costs or issues that are not included in the marginal cost to the firm? Or some social gains that are not included in what people pay for a good?

Introduction to a Monopoly
class="introduction"
Eskom's Medupi Power Station

Medupi power station at Lephalale, Limpopo, South Africa (Credit: By JMK - Own work, CC BY-SA 3.0)

"Shed Eskom's monopoly to solve the electricity crisis"

The following are extracts from the Free Market Foundation's Eustace Davie (2008) commentary on measures to solve South Africa's electricity crisis:

"There is no reason for the country to go through years of blackouts and electricity rationing, turning away foreign investment, and curtailing local capital expansion programmes. South Africa could have electricity without blackouts within a short time if government forthwith abandoned Eskom's monopoly and allowed competing electricity generators and distributors into the market..." (Davie: 2008)

"According to the World Bank, whenever states own and operate infrastructure, four institutional problems appear repeatedly. First, there is a misallocation of resources: a tendency to become involved in large-scale projects that are not economically viable. An example of this, of course, is Eskom producing generating capacity that far exceeded potential demand, as it did some years ago, leading to the 'moth-balling' / scrapping of high-cost generation plants that were redundant. Or, again just as Eskom has done, the state industry goes to the other extreme and fails to increase capacity to meet demand, causing huge disruption because there are no alternative suppliers, thanks to its legislated monopoly". (Davie: 2008)

"The second institutional problem is inadequate maintenance. An examination of the 'moth-balled' generation plants is likely to reveal that they were totally neglected, making reactivation impossible, while the records of existing plants most probably will reveal that earlier adequate maintenance would have avoided some of the current problems. A typical example of higher costs caused by neglect, mentioned by the World Bank in its 1994 Development Report, is (that the expenditure of) $12 billion in road maintenance in Africa would have saved $45 billion in road reconstruction". (Davie: 2008)

"A third problem is waste and inefficiency in the operation of infrastructure. For example, port facilities in developing countries move cargo from ship to shore at only 40 per cent the speed of the world's most efficient ports". (Davie: 2008)

"A fourth problem in the operation of much state-owned infrastructure is the lack of a sensible relationship between prices and costs. Prices are set politically, with the result that electric power prices of developing countries are set typically at half their cost, and the negative effects of over-usage are ignored. As Andrew Etzinger, Eskom's chief of demand-side management said in a recent media interview, 'The fact is, in this country, for a long time we have had a surplus at a cheap price – far cheaper than in other industrial nations. So it has made sense for the giant investors, whose plant needs massive amounts of electricity, to invest here.' What he did not add was that these benefits were provided at the expense of past, current and future taxpayers". (Davie: 2008)

"Nobody knows what electricity 'should cost' in South Africa. A real price could be determined only through freedom of entry into the generation, transmission and distribution of electricity, unfettered competition between the various providers, and the removal of all special privileges currently enjoyed by Eskom. In other words, Eskom would have to be broken up into separate functioning parts, and the ownership of those parts preferably handed over to the citizens of South Africa, to whom we are told they belong. These new private companies could then compete with, or be purchased by, the worlds' most efficient energy companies". (Davie: 2008)

"South Africans will derive a great deal more benefit from efficiently, competitively and reliably provided electricity than from the illusion that, as citizens, they 'own' and 'control' Eskom, especially if they, and most importantly the poor, have had the benefit of a payout from the sale of its assets" (Davie: 2008).

Note:
Introduction to a Monopoly
In this chapter, you will learn about:

- How monopolies form: barriers to entry
- How a profit-maximizing monopoly chooses output and price

There is a widespread belief that top executives at firms are the strongest supporters of market competition, but this belief is far from the truth. Think about it this way: If you very much wanted to win an Olympic gold medal, would you rather be far better than everyone else, or locked in competition with many athletes just as good as you are? Similarly, if you would like to attain a very high level of profits, would you rather manage a business with little or no competition, or struggle against many tough competitors who are trying to sell to your customers? By now, you might have read the chapter on Perfect Competition. In this chapter, we explore the opposite extreme: monopoly.

If perfect competition is a market where firms have no market power and they simply respond to the market price, monopoly is a market with no competition at all, and firms have complete market power. In the case of **monopoly**, one firm produces all of the output in a market. Since a monopoly faces no significant competition, it can charge any price it wishes. While a monopoly, by definition, refers to a single firm, in practice the term is often used to describe a market in which one firm merely has a very high market share.

Even though there are very few true monopolies in existence, we do deal with some of those few every day, often without realizing it: ESKOM, TELKOM, the South African Post Office and municipal refuse collection are a few examples. Some monopolies such as the state owned and operated utilities of ESKOM, TELKOM and the South African Post Office, for example, are artificially sustained by means of legislation limiting the providers of electricity, telecommunication services and postal service to certain providers. Other monopolies may occur more naturally due to exclusive access to resources or technology. For example some new medicines are produced by only one pharmaceutical firm—and no close substitutes for that drug may exist.

From the mid-1990s until 2004, the United States Department of Justice prosecuted the Microsoft Corporation for including Internet Explorer as the default web browser with its operating system. The Justice Department’s argument was that, since Microsoft possessed an extremely high market share in the industry for operating systems, the inclusion of a free web

browser constituted unfair competition to other browsers, such as Netscape Navigator. Since nearly everyone was using Windows, including Internet Explorer eliminated the incentive for consumers to explore other browsers and made it impossible for competitors to gain a foothold in the market. In 2013, the Windows system ran on more than 90% of the most commonly sold personal computers. In 2015, a United States federal court threw out anti-competitive behavior (monopoly) charges that Google had an agreement with mobile device makers to set Google as the default search engine.

South Africa's Competition Commission and Competition Tribunal also monitor and prosecute cases of anti-competitive behaviour by local firms. Check out this link which lists South Africa's Competition Commission Tribunal Decisions to date.

```
Competition Commission Tribunal Decisions
  [link].
```

This chapter begins by describing how monopolies are protected from competition, including laws that prohibit competition, technological advantages, and certain types of markets. It then discusses how a monopoly will choose its profit-maximizing quantity to produce and what price to charge. While a monopoly must be concerned about whether consumers will purchase its products or spend their money on something altogether different, the monopolist need not worry about the actions of other competing firms producing its products. As a result, a monopoly is not a price taker like a perfectly competitive firm, but instead exercises some power to choose its market price.

References:

Crisis meetings to end Eskom chaos. 2008. iol news/South Africa 20 January 09:02am by Edwin Naidu, Justin Brown, Eleanor Momberg, Mary Corrigall, Keith Ross and Jeremy Gordin Available http://www.iol.co.za/news/south-africa/crisis-meetings-to-end-eskom-chaos-386324 (Accessed: 14 May 2016)

Davie, E. 2008. Shed Eskom’s monopoly to solve the electricity crisis. Moneyweb, 27 January. Available

http://www.moneyweb.co.za/archive/shed-eskoms-monopoly-to-solve-the-electricity-cris/ (Accessed: 14 May 2016)

World Bank. 1994. The World Bank annual report 1994. Washington DC ; World Bank. Available http://documents.worldbank.org/curated/en/1994/08/698118/world-bank-annual-report-1994 (Accessed: 14 May 2016)

How Monopolies Form: Barriers to Entry

By the end of this section, you will be able to:

- Distinguish between a natural monopoly and a legal monopoly.
- Explain how economies of scale and the control of natural resources led to the necessary formation of legal monopolies
- Analyze the importance of trademarks and patents in promoting innovation
- Identify examples of predatory pricing

Figure 1.

Because of the lack of competition, monopolies tend to earn significant economic profits. These profits should attract healthy competition as

described in Perfect Competition, and yet, because of one particular characteristic of monopoly, they do not. **Barriers to entry** are the legal, technological, or market forces that discourage or prevent potential competitors from entering a market. Barriers to entry can range from the simple and easily surmountable, such as the cost of renting retail space, to the extremely restrictive. For example, there are a finite number of radio frequencies available for broadcasting. Once the rights to all of them have been purchased, no new competitors can enter the market.

In some cases, barriers to entry may lead to monopoly. In other cases, they may limit competition to a few firms. Barriers may block entry even if the firm or firms currently in the market are earning profits. Thus, in markets with significant barriers to entry, it is *not* true that abnormally high profits will attract new firms, and that this entry of new firms will eventually cause the price to decline so that surviving firms earn only a normal level of profit in the long run.

There are two types of monopoly, based on the types of barriers to entry they exploit. One is **natural monopoly**, where the barriers to entry are something other than legal prohibition (man-made or artificial limitations). The other is **legal (man-made or artificial) monopoly**, where laws prohibit (or severely limit) competition.

Natural Monopoly

Economies of scale can combine with the size of the market to limit competition. Figure 2 presents a long-run average cost curve for the airplane manufacturing industry. It shows economies of scale up to an output of 8,000 planes per year at a unit cost of C0, then constant returns to scale from 8,000 to 20,000 planes per year, and diseconomies of scale at a quantity of production greater than 20,000 planes per year.

For a firm to maximise its efficiency in this market it needs to produce at least 8,000 planes/year.Let's say there is already one firm (a monopoly) operating in this market and it is indeed manufacturing 8000 planes/year. If a second firm attempts to enter the market at a smaller size, say by producing a quantity of about 5,000 planes, then its average costs will be

higher at C1 than that of the existing firm (C0), and it will be unable to compete. If the second firm attempts to enter the market at a larger size, like 8,000 planes per year, then it could also produce at a lower average cost—but it probably could not sell all 8,000 planes that it produced because of insufficient demand in the market.

Economies of Scale and Natural Monopoly

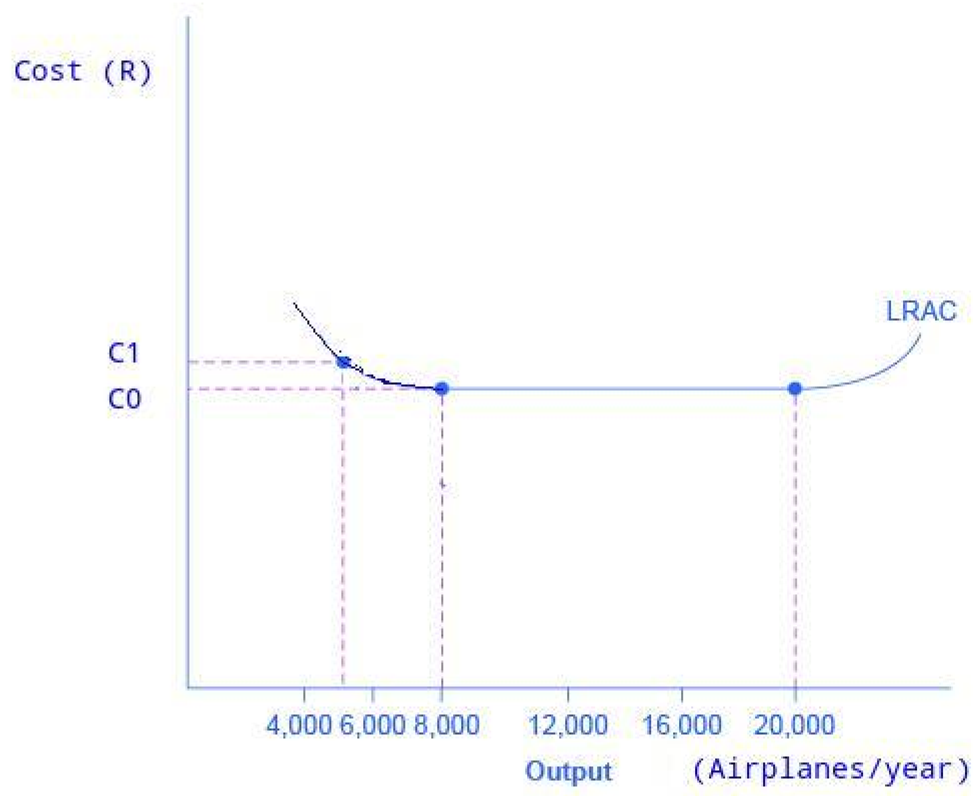


Figure 2: A natural monopoly occurs when the quantity demanded is less than the minimum quantity it takes to be at the bottom of the long-run average cost curve.

This situation, when economies of scale are large relative to the quantity demanded in the market, is called a natural monopoly. Natural monopolies often arise in industries where the marginal cost of adding an additional customer is very low, once the fixed costs of the overall system are in place. Once the main water pipes are laid through a neighborhood, the marginal cost of providing water service to another home is fairly low. Once electricity lines are installed through a neighborhood, the marginal cost of

providing additional electrical service to one more home is very low. It would be costly and duplicative for a second water company to enter the market and invest in a whole second set of main water pipes, or for a second electricity company to enter the market and invest in a whole new set of electrical wires. These industries offer an example where, because of economies of scale, one producer can serve the entire market more efficiently than a number of smaller producers that would need to make duplicate physical capital investments.

A natural monopoly can also arise in smaller local markets for products that are difficult to transport. For example, cement production exhibits economies of scale, and the quantity of cement demanded in a local area may not be much larger than what a single plant can produce. Moreover, the costs of transporting cement over land are high, and so a cement plant in an area without access to water transportation may be a natural monopoly.

Control of a Physical Resource

Another type of natural monopoly occurs when a company has control of a scarce physical resource. In the South African economy, one historical example of this pattern is the De Beers diamond company.In the first half of the 20th century, De Beers secured complete control of world diamond supplies through its distribution channel, the Diamond Trading Company (Zimnisky, 2013). This dominant position, secured by seizing control of the supply of diamonds, allowed it to control the prices of these precious stones. Today De Beers no longer has the same degree of control over diamond supplies as there are more competitors.

Legal Monopoly

For some products, the government erects barriers to entry by prohibiting or limiting competition. Under South African law, for example, no organization except the South African Postal Service is legally allowed to deliver first-class mail. Similarly, many municipalities have laws or regulations that allow households no choice of electricity or water supply. These have to be purchased from the local authority. Municipalities also reserve for themselves the sole right to supply sanitation/sewage and refuse

removal services. Most legal monopolies are considered utilities—products necessary for everyday life—that are socially beneficial to have. As a consequence, the government allows producers to become regulated monopolies, to insure that an appropriate amount of these products is provided to consumers. Additionally, legal monopolies are often subject to economies of scale, so it makes sense to allow only one provider.

Promoting Innovation

Innovation takes time and resources to achieve. Suppose a company invests in research and development and finds the cure for the common cold. In a world in which information is increasingly freely available (especially through the internet), other companies could take the formula, produce the drug, and because they did not incur the costs of research and development (R and D), undercut the price of the company that discovered the drug. Given this possibility, many firms would choose not to invest in research and development, and as a result, the world would have less innovation and fewer new products. To prevent this from happening, countries promulgate patent and copyright law to protect the intellectual property (ideas) of individuals.

South Africa's Companies and Intellectual Property Commission (CIPC) is the custodian of all new patent applications that are filed within the Republic of South Africa in terms of the South African Patent Act 57 of 1978. A **patent** gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time; in South Africa, exclusive patent rights last for 20 years. The idea is to provide limited monopoly power so that innovative firms can recoup their investment in research and development (R and D), but then to allow other firms to produce the product more cheaply once the patent expires.

A **trademark** is an identifying symbol or name for a particular good, like the Toyota badge on its cars, or the Nike "swoosh" that appears on shoes and athletic gear. In South Africa a trade mark can be protected and defended under the Trade Marks Act, 1993 (Act 194 of 1993) if it is registered with the CIPC. Within South Africa a registered trade mark can

be protected forever, provided it is renewed every ten (10) years upon payment of the prescribed renewal fee.

A **copyright**, according to South Africa's Copyright Act of 1978 refers to the right to control the use and distribution of artistic and creative works. No one can reproduce, display, or perform a copyrighted work without permission of the author. Copyright protection in South Africa ordinarily lasts for the life of the author plus 50 years.

Roughly speaking, patent law covers inventions and copyright protects books, songs, and art. But in certain areas, like the invention of new software, it has been unclear whether patent or copyright protection should apply. There is also a body of law known as **trade secrets**. Even if a company does not have a patent on an invention, competing firms are not allowed to steal their secrets. One famous trade secret is the formula for Coca-Cola, which is not protected under copyright or patent law, but is simply kept secret by the company.

Taken together, this combination of patents, trademarks, copyrights, and trade secret law is called **intellectual property**, because it implies ownership over an idea, concept, or image, not a physical piece of property like a house or a car. Countries around the world have enacted laws to protect intellectual property, although the time periods and exact provisions of such laws vary across countries. There are ongoing negotiations, both through the World Intellectual Property Organization (WIPO) and through international treaties, to bring greater harmony between the intellectual property laws of different countries and to determine the extent to which patents and copyrights of one country will be respected in other countries.

The combination of improvements in production technologies and a general sense that the markets could provide services adequately led to a wave of **deregulation**, starting in the late 1970s and continuing into the 1990s. This wave eliminated or reduced government restrictions on the firms that could enter, the prices that could be charged, and the quantities that could be produced in many industries, including telecommunications, airlines, trucking, banking, and electricity.

Around the world, from Europe to Latin America to Africa and Asia, many governments continue to control and limit competition in what those governments perceive to be key industries, including airlines, banks, steel companies, oil companies, and telephone companies.

Note:
Vist this website for examples of some pretty strange patents.

Intimidating Potential Competition

Businesses have developed a number of schemes for creating barriers to entry by deterring potential competitors from entering the market. One method is known as **predatory pricing**, in which a firm uses the threat of sharp price cuts to discourage competition. Predatory pricing, along with price discrimination, is a violation of South Africa's Competition Act 89 of 1998, but it is difficult to prove.

Consider a large airline that provides most of the flights between two particular cities. A new, small start-up airline decides to offer service between these two cities. The large airline immediately slashes prices on this route to the bone, so that the new entrant cannot make any money. After the new entrant has gone out of business, the large airline then raises prices again.

After this pattern is repeated once or twice, potential new entrants may decide that it is not wise to try to compete. Small airlines often accuse

larger airlines of predatory pricing and other anti-competitive behavior such as making exclusionary agreements with agents. In 2006, for example, South African Airways (SAA) reached a settlement with the Competition Commission about allegations (made by rivals Comair and Nationwide airlines) of entering into agreements with travel agents to deal only with South African Airways (Jenkins et al: 2009). South African Airways was fined R15 million for its infringement of the Competition Act.

In some cases, large advertising budgets can also act as a way of discouraging the competition. If the only way to launch a successful new national cola drink is to spend more than the promotional budgets of Coca-Cola and Pepsi Cola, not too many companies will try. A firmly established brand name can be difficult to dislodge.

Summing Up Barriers to Entry

Table 1 lists the barriers to entry that have been discussed here. This list is not exhaustive, since firms have proved to be highly creative in inventing business practices that discourage competition. When barriers to entry exist, perfect competition is no longer a reasonable description of how an industry works. When barriers to entry are high enough, monopoly can result.

Barrier to Entry	**Government Role?**	**Example**
Natural monopoly	Government often responds with regulation (or ownership)	Water and electric companies

Barrier to Entry	Government Role?	Example
Control of a physical resource	No	DeBeers for diamonds
Legal monopoly	Yes	Post office, past regulation of airlines and telecommunications
Patent, trademark, and copyright	Yes, through protection of intellectual property	New medicines or software
Intimidating potential competitors	Somewhat	Predatory pricing; well-known brand names;exclusionary agreements

Barriers to Entry

Key Concepts and Summary

Barriers to entry prevent or discourage competitors from entering the market. These barriers include: economies of scale that lead to natural monopoly; control of a physical resource; legal restrictions on competition; patent, trademark and copyright protection; and practices to intimidate the competition like predatory pricing. Intellectual property refers to legally guaranteed ownership of an idea, rather than a physical item. The laws that protect intellectual property include patents, copyrights, trademarks, and trade secrets. A natural monopoly arises when economies of scale persist over a large enough range of output that if one firm supplies the entire market, no other firm can enter without facing a cost disadvantage.

Self-Check Questions

Exercise:

Problem:

Examine each of the following cases and say whether they are: a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or, a situation that does not involve a barrier to entry.

a. A patented invention
b. A popular but easily copied restaurant recipe
c. An industry where economies of scale are very small compared to the size of demand in the market
d. A well-established reputation for slashing prices in response to new entry
e. A well-respected brand name that has been carefully built up over many years

---

Solution:

a. A patent is a government-enforced barrier to entry.
b. This is not a barrier to entry.
c. This is not a barrier to entry.
d. This is a barrier to entry, but it is not government-enforced.
e. This is a barrier to entry, but it is not directly government enforced.

Exercise:

Problem:

Examine each of the following cases and say whether they are: a government-enforced barrier to entry, a barrier to entry that is not government-enforced, or, a situation that does not involve a barrier to entry.

a. A city passes a law on how many licenses it will issue for taxicabs
b. A city passes a law that all taxicab drivers must pass a driving safety test and have insurance
c. A well-known trademark
d. Owning a spring that offers very pure water
e. An industry where economies of scale are very large compared to the size of demand in the market

---

Solution:

a. This is a government-enforced barrier to entry.
b. This is an example of a government law, but perhaps it is not much of a barrier to entry if most people can pass the safety test and get insurance.
c. Trademarks are enforced by government, and therefore are a barrier to entry.
d. This is probably not a barrier to entry, since there are a number of different ways of getting pure water.
e. This is a barrier to entry, but it is not government-enforced.

Exercise:

Problem:

Suppose the local electrical utility (Eskom), a legal monopoly based on economies of scale, was split into four firms of equal size, with the idea that eliminating the monopoly would promote competitive pricing of electricity. What do you anticipate would happen to prices?

Solution:

Because of economies of scale, each firm would produce at a higher average cost than before. (They would each have to build their own power lines.) As a result, they would each have to raise prices to cover their higher costs. The policy would fail.

Exercise:

Problem:

If government reduced the period of patent protection from 20 years to 10 years, what would likely happen to the amount of private research and development?

Solution:

Shorter patent protection would make innovation less lucrative, so the amount of research and development would likely decline.

Review Questions

Exercise:

Problem: How is monopoly different from perfect competition?

Exercise:

Problem: What is a barrier to entry? Give some examples.

Exercise:

Problem: What is a natural monopoly?

Exercise:

Problem: What is a legal (or artificial/man-made) monopoly?

Exercise:

Problem: What is predatory pricing?

Exercise:

Problem: How is intellectual property different from other property?

Exercise:

Problem:

By what legal mechanisms is intellectual property protected?

Exercise:

Problem: In what sense is a natural monopoly “natural”?

Critical Thinking Questions

Exercise:

Problem:

DeBeers Diamonds does not have the monopoly power it once had. How do you suppose its barriers to entry were weakened?

Exercise:

Problem:

Why are generic ("no-name brand") pharmaceuticals significantly cheaper than name brand ones?

Exercise:

Problem:

For many years, the United States Justice Department has tried to break up large firms like IBM, Microsoft, and most recently Google, on the grounds that their large market share made them essentially monopolies. In a global market, where United States firms compete with firms from other countries, would this policy make the same sense as it might in a purely domestic context (in the domestic United States market)?

Exercise:

Problem:

Intellectual property laws are intended to promote innovation but some economists have argued that such laws are not desirable. What do you suppose is the basis for their reasoning?

References

Jenkins, H., Niels, G. and Noble, R. 2009. The South African Airways cases: blazing a trail for Europe to follow? Prepared for the Third Annual Competition Conference, Pretoria, August 14 . Available http://www.compcom.co.za/wp-content/uploads/2014/09/The-South-African-Airways-cases.pdf (Accessed: 14 May 2016)

Zimnisky, P. 2013. “A Diamond Market No Longer Controlled By De Beers.” Available http://www.kitco.com/ind/Zimnisky/2013-06-06-A-Diamond-Market-No-Longer-Controlled-By-De-Beers.html. (Accessed: 3 May 2016)

Glossary

barriers to entry
: the legal, technological, or market forces that may discourage or prevent potential competitors from entering a market

copyright
: a form of legal protection to prevent copying, for commercial purposes, original works of authorship, including books and music

deregulation
: removing government controls over setting prices and quantities in certain industries

intellectual property
: the body of law including patents, trademarks, copyrights, and trade secret law that protect the right of inventors to produce and sell their

inventions

legal monopoly
legal prohibitions against competition, such as regulated monopolies and intellectual property protection

monopoly
a situation in which one firm produces all of the output in a market

natural monopoly
economic conditions in the industry, for example, economies of scale or control of a critical resource, that limit effective competition

patent
a government rule that gives the inventor the exclusive legal right to make, use, or sell the invention for a limited time

predatory pricing
when an existing firm uses sharp but temporary price cuts to discourage new competition

trade secrets
methods of production kept secret by the producing firm

trademark
an identifying symbol or name for a particular good and can only be used by the firm that registered that trademark

How a Profit-Maximizing Monopoly Chooses Output and Price

By the end of this section, you will be able to:

- Explain the perceived demand curve for a perfect competitor and a monopoly
- Analyze a demand curve for a monopoly and determine the output that maximizes profit and revenue
- Calculate marginal revenue and marginal cost
- Explain allocative efficiency as it pertains to the efficiency of a monopoly

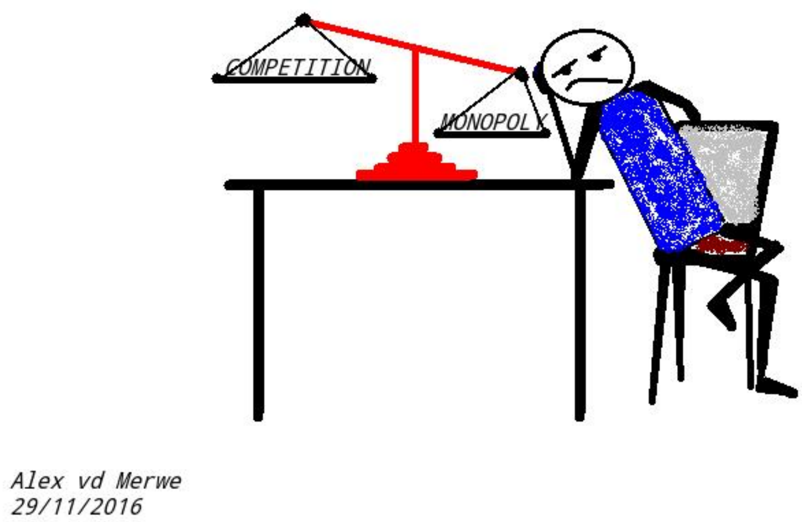


Figure 1. Perfect competition versus monopoly.

Consider a monopoly firm, comfortably surrounded by barriers to entry so that it need not fear competition from other producers. How will this monopoly choose its profit-maximizing quantity of output, and what price will it charge? Profits for the monopolist, like any firm, will be equal to total revenues minus total costs. The pattern of costs for the monopoly can be analyzed within the same framework as the costs of a perfectly competitive firm—that is, by using total cost, fixed cost, variable cost, marginal cost, average cost, and average variable cost. However, because a monopoly faces no competition, its situation and its decision process will differ from that of a perfectly competitive firm. (The Clear it Up feature discusses how hard it is sometimes to define "market" in a monopoly situation.)

Demand Curves Perceived by a Perfectly Competitive Firm and by a Monopoly

A perfectly competitive firm acts as a price taker, so its calculation of total revenue is made by taking the given market price and multiplying it by the quantity of output that the firm chooses. The demand curve *as it is perceived by a perfectly competitive firm* appears in Figure 2 (a). The flat perceived demand curve means that, from the viewpoint of the perfectly competitive firm, it could sell either a relatively low quantity like Ql or a relatively high quantity like Qh at the market price P.

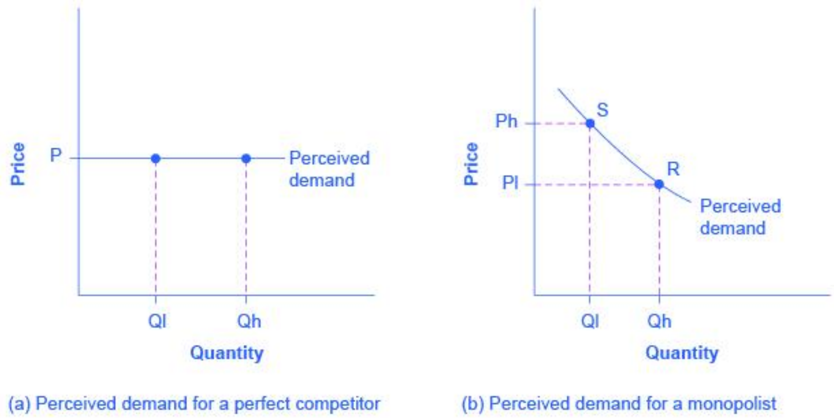


Figure 2: (a) A perfectly competitive firm perceives the demand curve that it faces to be flat. The flat shape means that the firm can sell either a low quantity (Ql) or a high quantity (Qh) at exactly the same price (P). (b) A monopolist perceives the demand curve that it faces to be the same as the market demand curve, which for most goods is downward-sloping. Thus, if the monopolist chooses a high level of output (Qh), it can charge only a relatively low price (Pl); conversely, if the monopolist chooses a low level of output (Ql), it can then charge a higher price (Ph). The challenge for the monopolist is to choose the combination of price and quantity that maximizes profits.

Note:

What defines the market?

A monopoly is a firm that sells all or nearly all of the goods and services in a given market. But what defines the “market”?

Buyers and sellers make up a market. In some cases there are many buyers and sellers (perfect competition) and in others there may be only one (monopoly) or a few sellers (oligopoly) and only one buyer (monopsony) or only a few buyers (oligopsony). Given different market conditions sellers and buyers behave differently. In a famous 1947 case, the federal government accused the DuPont company of having a monopoly in the cellophane market, pointing out that DuPont produced 75% of the cellophane in the United States. DuPont countered that even though it had a 75% market share in cellophane, it had less than a 20% share of the “flexible packaging materials,” which includes all other moisture-proof papers, films, and foils. In 1956, after years of legal appeals, the United States Supreme Court held that the broader market definition was more appropriate, and the case against DuPont was dismissed.

In 2014 the South African Competition Tribunal found Sasol Chemical Industries Limited (“SCI”), which is a subsidiary of Sasol Ltd (“Sasol”), guilty of charging domestic customers excessive prices for purified propylene and polypropylene (used to make plastics) between January 2004 and December 2007 (Competition Tribunal, 2014). The Tribunal argued that the price SCI charged Safripol, SCI’s only external customer for purified propylene and a competitor of SCI in other areas of production, was to Safripol’s detriment and inhibited its ability to effectively compete with SCI. It was also of the view that, SCI’s locally charged polypropylene prices have had a significant adverse effect on the local plastic converters and caused them harm during the complaint period. Sasol was duly fined in excess of R500 million for this infringement. (Competition Tribunal South Africa: 2014)

Questions over how to define the market continue today not only in South Africa but also globally. Microsoft in the 1990s had a dominant share of the software for computer operating systems, but in the total market for all computer software and services, including everything from games to scientific programs, the Microsoft share was only about 14% in 2014. De Beers has a monopoly in diamonds, but it has a much smaller share of

the total market for precious gemstones and an even smaller share of the total market for jewelry. A small town like Mooiriver in the KwaZulu-Natal Midlands may have only one petrol station: is this fuel station a “monopoly,” or does it compete with fuel stations that might be five, 10, or 50 km away?

In general, if a firm produces a product without close substitutes, then the firm can be considered a monopoly producer in a single market. But if buyers have a range of similar—even if not identical—options available from other firms, then the firm is not a monopoly. Still, arguments over whether substitutes are close or not close can be controversial.

While a monopolist can charge *any* price for its product, that price is nonetheless constrained by demand for the firm’s product. No monopolist, even one that is thoroughly protected by high barriers to entry, can require consumers to purchase its product. Because the monopolist is the only firm in the market, its demand curve is the same as the market demand curve, which is, unlike that for a perfectly competitive firm, downward-sloping.

Figure 2 illustrates this situation. The monopolist can either choose a point like R with a low price (Pl) and high quantity (Qh), or a point like S with a high price (Ph) and a low quantity (Ql), or some intermediate point. Setting the price too high will result in a low quantity sold, and will not bring in much revenue. Conversely, setting the price too low may result in a high quantity sold, but because of the low price, it will not bring in much revenue either. The challenge for the monopolist is to strike a profit-maximizing balance between the price it charges and the quantity that it sells. But why isn’t the perfectly competitive firm’s demand curve also the market demand curve? See the following Clear it Up feature for the answer to this question.

Note:

What is the difference between perceived demand and market demand?

The demand curve as perceived by a perfectly competitive firm is not the overall market demand curve for that product. However, the firm’s demand curve as perceived by a monopoly is the same as the market demand curve. The reason for the difference is that each perfectly competitive firm perceives the demand for its products in a market that includes many other firms; in effect, the demand curve perceived by a perfectly competitive firm is a tiny slice of the entire market demand curve. In contrast, a monopoly perceives demand for its product in a market where the monopoly is the only producer.

Total Cost and Total Revenue for a Monopolist

Profits for a monopolist can be illustrated with a graph of total revenues and total costs, as shown with the example of the hypothetical HealthPill firm in Figure 3. The total cost curve has its typical shape; that is, total costs rise and the curve grows steeper as output increases.

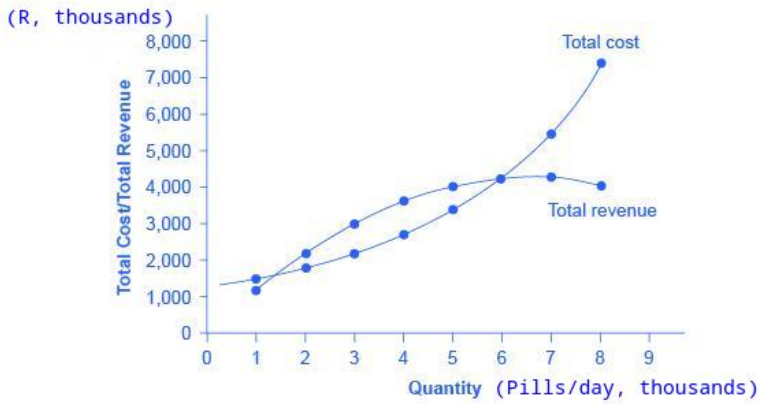


Figure 3: Total revenue for the monopoly firm called HealthPill first rises, then falls. Low levels of output bring in relatively little total revenue, because the quantity is low. High levels of output bring in relatively less revenue, because the high quantity pushes down the market price. The total cost curve is upward-sloping. Profits will be highest at the quantity of output where total revenue is most above total cost. Of the choices in Table 1, the highest profits happen at an output of 4000 pills/day. The profit-maximizing level of output is not the same as the revenue-maximizing level of output, which should make sense, because profits take costs into account and revenues do not.

Quantity (Pills/day)	**Total Cost**	**Quantity**	**Price**	**Total Revenue**	**Profit = Total Revenue – Total Cost**
1,000	1,500 000	1,000	1,200	1,200 000	–300 000
2,000	1,800 000	2,000	1,100	2,200 000	400 000
3,000	2,200 000	3, 000	1,000	3,000 000	800 000
4,000	2,800 000	4, 000	900	3,600 000	800 000
5,000	3,500 000	5, 000	800	4,000 000	500 000
6,000	4,200 000	6,000	700	4,200 000	0
7,000	5,600 000	7, 000	600	4,200 000	–1,400 000

Quantity (Pills/day)	Total Cost	Quantity	Price	Total Revenue	Profit = Total Revenue – Total Cost
8,000	7,400 000	8, 000	500	4,000 000	–3,400 000

Total Costs and Total Revenues of HealthPill (Rands)

To calculate total revenue for a monopolist, start with the demand curve perceived by the monopolist. Table 1 shows quantities along the demand curve and the price at each quantity demanded, and then calculates total revenue by multiplying price times quantity at each level of output. As the figure illustrates, total revenue for a monopolist rises, flattens out, and then falls. In this example, total revenue is highest at a quantity of 6,000 or 7,000 pills per day.

Clearly, the total revenue for a monopolist is not a straight upward-sloping line, in the way that total revenue was for a perfectly competitive firm. The different total revenue pattern for a monopolist occurs because the quantity that a monopolist chooses to produce affects the market price, which was not true for a perfectly competitive firm. If the monopolist charges a very high price, then quantity demanded drops, and so total revenue is very low. If the monopolist charges a very low price, then, even if quantity demanded is very high, total revenue will not add up to much. At some intermediate level, total revenue will be highest.

However, the monopolist is not seeking to maximize revenue, but instead to earn the highest possible profit. Profits are calculated in the final row of the table. In the HealthPill example in Figure 2, the highest profit will occur at the quantity where total revenue is the farthest above total cost. Of the choices given in the table, the highest profits occur at an output of 4,000 pills per day where total profit is R800 000.

Marginal Revenue and Marginal Cost for a Monopolist

In the real world, a monopolist often does not have enough information to analyze its entire total revenues or total costs curves; after all, the firm does not know exactly what would happen if it were to alter production dramatically. But a monopolist often has fairly reliable information about how changing output by small or moderate amounts will affect its marginal revenues and marginal costs, because it has had experience with such changes over time and because the effects of small changes are easier to figure out from current experience. A monopolist, like a perfectly competitive firm, can use information on marginal revenue and marginal cost to seek out the profit-maximizing combination of quantity and price.

The first four columns of Table 2 use the numbers on total cost from the HealthPill example in the previous table and calculate marginal cost and average cost. This monopoly faces a typical upward-sloping marginal cost curve, as shown in Figure 4. The second four columns of Table 2 also use the total revenue information from Table 1 and calculate marginal revenue.

Notice that marginal revenue is zero at a quantity of 7, and turns negative at quantities higher than 7. It may seem unbelievable that marginal revenue could ever be zero or negative: after all, does an increase in quantity sold not always mean more revenue? For a perfect competitor, each additional unit sold brought a positive marginal revenue, because marginal revenue was equal to the given market price. But a monopolist can sell a larger quantity and see a decline in total revenue. When a monopolist increases sales by one unit, it gains some marginal revenue from selling that extra unit, but also loses some marginal revenue because every other unit must now be sold at a lower price. As the quantity sold becomes higher, the drop in price affects a greater quantity of sales, eventually causing a situation where more sales cause marginal revenue to be negative.

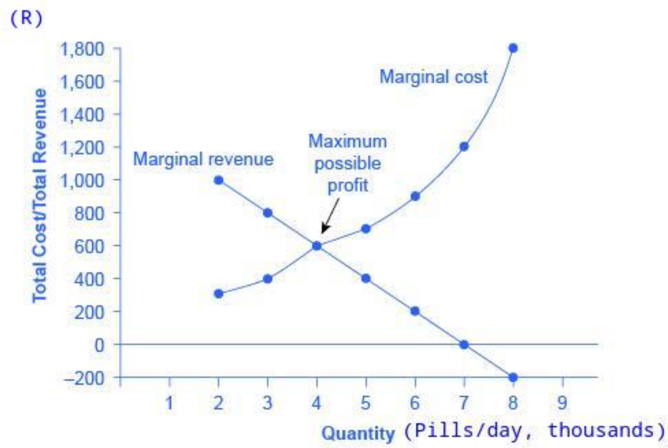


Figure 4: For a monopoly like HealthPill, marginal revenue decreases as additional units are sold. The marginal cost curve is upward-sloping. The profit-maximizing choice for the monopoly will be to produce at the quantity where marginal revenue is the same as the marginal cost: that is, MR = MC. If the monopoly produces a lower quantity, then MR is less than MC at those levels of output, and the firm can make higher profits by expanding output. If the firm produces at a greater quantity, then MC exceeds MR, and the firm can make higher profits by reducing its quantity of output.

Cost Information				**Revenue Information**			
Quantity (Pills/day)	**Total Cost**	**Marginal Cost**	**Average Cost**	**Quantity**	**Price**	**Total Revenue**	**Marginal Revenue**
1,000	1,500 000	1,500	1,500 000	1,000	1,200	1,200 000	1,200
2,000	1,800 000	300	900 000	2,000	1,100	2,200 000	1,000
3,000	2,200 000	400	733 000	3,000	1,000	3,000 000	800
4,000	2,800 000	600	700 000	4,000	900	3,600 000	600
5,000	3,500 000	700	700 000	5,000	800	4,000 000	400
6,000	4,200 000	700	700 000	6,000	700	4,200 000	200

Cost Information				Revenue Information			
7,000	5,600 000	1,400	800 000	7,000	600	4,200 000	0
8,000	7,400 000	1,800	925 000	8,000	500	4,000 000	–200

Costs and Revenues of HealthPill (Rands)

A monopolist can determine its profit-maximizing price and quantity by analyzing the marginal revenue and marginal costs of producing an extra unit. If the marginal revenue exceeds the marginal cost, then the firm should produce the extra unit.

For example, at an output of 3,000 pills per day in Figure 4, marginal revenue is R800 and marginal cost is R400 so producing these extra 1000 units will clearly add to overall profits. At an output of 4,000 pills per day marginal revenue is R600 and marginal cost is R600 so producing this volume of pills still means overall profits are unchanged. However, expanding output from 4,000 to 5,000 pills per day would involve a marginal revenue of R400 and a marginal cost of R700 so that the additional 1,000 units beyond 4000 pills per day would actually reduce profits. Thus, the monopolist can tell from the marginal revenue and marginal cost that, of the choices given in the table, the profit-maximizing level of output is 4,000 pills per day.

Indeed, the monopoly could seek out the profit-maximizing level of output by increasing quantity by a small amount, calculating marginal revenue and marginal cost, and then either increasing output as long as marginal revenue exceeds marginal cost or reducing output if marginal cost exceeds marginal revenue. This process works without any need to calculate total revenue and total cost. Thus, a profit-maximizing monopoly should follow the rule of producing up to the quantity where marginal revenue is equal to marginal cost—that is, MR = MC. Well, well...this is just the same as for a perfectly competitive firm!

Note:
Maximizing Profits
If you find it does not make sense that producing where marginal revenue equals marginal cost will maximize profits, working through the numbers will help.
Step 1. Remember that marginal cost is defined as the change in total cost from producing a small amount of additional output.
Equation:

$$\text{MC} = \frac{\text{change in total cost}}{\text{change in quantity produced}}$$

Step 2. Note that in Table 2, as output increases from 1,000 to 2,000 units, total cost increases from R1,500 000 to R1,800 000. As a result, the marginal cost of the second thousand units/day will be:
Equation:

$$\begin{aligned} \text{MC} &= \frac{\text{R1,800 000–R1,500 000}}{\text{1,000}} \\ &= \text{R300} \end{aligned}$$

Step 3. Remember that, similarly, marginal revenue is the change in total revenue from selling a small amount of additional output.
Equation:

$$\text{MR} = \frac{\text{change in total revenue}}{\text{change in quantity sold}}$$

Step 4. Note that in Table 2, as output increases from 1,000 to 2,000 units per day, total revenue increases from R1,200 000 to R2,200 000. As a result, the marginal revenue of the second unit will be:
Equation:

$$\begin{aligned} MR &= \frac{R2{,}200\ 000 - R1{,}200\ 000}{1{,}000} \\ &= R1000 \end{aligned}$$

Quantity (Pills/day)	Marginal Revenue	Marginal Cost	Marginal Profit	Total Profit
1,000	1,200	1,500	–300	–300
2,000	1,000	300	700	400
3,000	800	400	400	800
4,000	600	600	0	800
5,000	400	700	–300	500
6,000	200	700	–500	0
7,000	0	1,400	–1,400	–1,400

Marginal Revenue, Marginal Cost, Marginal and Total Profit

Table 3 repeats the marginal cost and marginal revenue data from Table 2, and adds two more columns: **Marginal profit** is the profitability of each additional unit sold. It is defined as marginal revenue minus marginal cost. Finally, total profit is the sum of marginal profits. As long as marginal profit is positive, producing more output will increase total profits. When marginal profit turns negative, producing more output will decrease total profits. Total profit is maximized where marginal revenue equals marginal cost. In this example, maximum profit occurs at 4,000 units of output per day.

A perfectly competitive firm will also find its profit-maximizing level of output where MR = MC. The key difference with a perfectly competitive firm is that in the case of perfect competition, marginal revenue is equal to price (MR = P), while for a monopolist, marginal revenue is not equal to the price, because changes in quantity of output affect the price.

Illustrating Monopoly Profits

It is straightforward to calculate profits of given numbers for total revenue and total cost. However, the size of monopoly profits can also be illustrated graphically with Figure 5, which takes the marginal cost and marginal revenue curves from the previous exhibit and adds an average cost curve and the monopolist's perceived demand curve.
Illustrating Profits at the HealthPill Monopoly

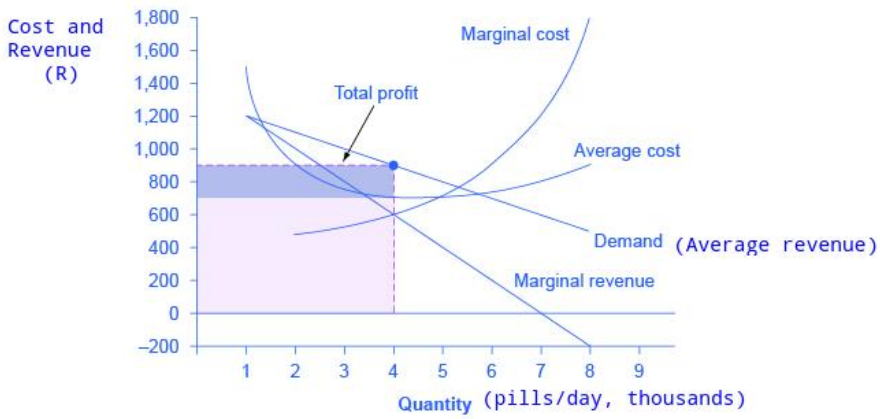


Figure 5: This figure begins with the same marginal revenue and marginal cost curves from the HealthPill monopoly presented in Figure 3. It then adds an average cost curve and the demand curve faced by the monopolist. The HealthPill firm first chooses the quantity where MR = MC; in this example, the quantity is 4,000 units/day. The monopolist then decides what price to charge by looking at the demand curve it faces. The large box, with quantity on the horizontal axis and marginal revenue on the vertical axis, shows total revenue for the firm. Total costs for the firm are shown by the lighter-shaded box, which is quantity on the horizontal axis and marginal cost of production on the vertical axis. The large total revenue box minus the smaller total cost box leaves the darkly shaded box that shows total profits. Since the price charged is above average cost, the firm is earning positive profits.

Figure 6 illustrates the three-step process where a monopolist: selects the profit-maximizing quantity to produce; decides what price to charge; determines total revenue, total cost, and profit.

Step 1: The Monopolist Determines Its Profit-Maximizing Level of Output

The firm can use the points on the demand curve D (also known as the average revenue curve) to calculate total revenue, and then, based on total revenue, calculate its marginal revenue curve. The profit-maximizing quantity will occur where MR = MC—or at the last possible point before marginal costs start exceeding marginal revenue. In Figure 4, MR = MC occurs at an output of 4,000 units per day.

Step 2: The Monopolist Decides What Price to Charge

The monopolist will charge what the market is willing to pay. A dotted line drawn straight up from the profit-maximizing quantity to the demand curve (or average revenue curve) shows the profit-maximizing price. This price is above the average cost curve, which shows that the firm is earning profits.

Step 3: Calculate Total Revenue, Total Cost, and Profit

Total revenue is the overall shaded box, where the width of the box is the quantity being sold and the height is the price. In Figure 5, the bottom part of the shaded box, which is shaded more lightly, shows total costs; that is, quantity on the horizontal axis multiplied by average cost on the vertical axis. The larger box of total revenues minus the smaller box of total costs will equal profits, which is shown by the darkly shaded box. In a perfectly competitive market, the forces of entry would erode this profit in the long run. But a monopolist is protected by barriers to entry. In fact, one telltale sign of a possible monopoly is when a firm earns profits year after year, while doing more or less the same thing, without ever seeing those profits eroded by increased competition.

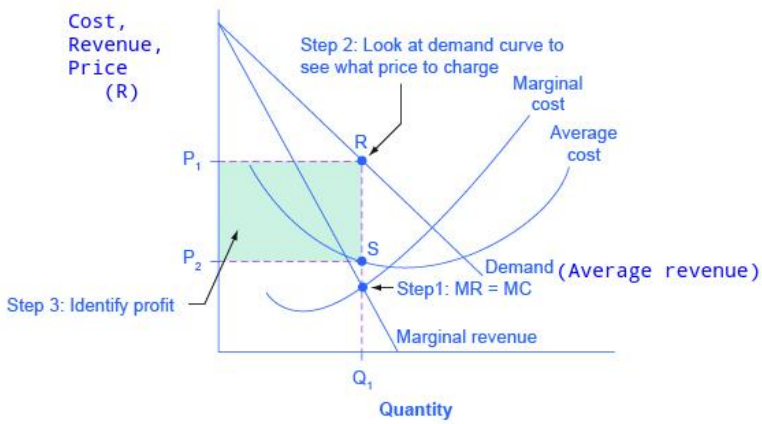


Figure 6: In Step 1, the monopoly chooses the profit-maximizing level of output Q1, by choosing the quantity where MR = MC. In Step 2, the monopoly decides how much to charge for output level Q1 by drawing a line straight up from Q1 to point R on its perceived demand curve. Thus, the monopoly will charge a price (P1). In Step 3, the monopoly identifies its profit. Total revenue will be Q1 multiplied by P1. Total cost will be Q1 multiplied by the average cost of producing Q1, which is shown by point S on the average cost curve to be P2. Profits will be the total revenue rectangle minus the total cost rectangle, shown by the shaded zone in the figure.

Note:

Why is a monopolist's marginal revenue always less than the price?

The marginal revenue curve for a monopolist always lies beneath the market demand curve. To understand why, think about increasing the quantity along the demand curve by one unit, so that you take one step down the demand curve to a slightly higher quantity but a slightly lower price. A demand curve is not sequential: It is not that first we sell Q_1 at a higher price, and then we sell Q_2 at a lower price. Rather, a demand curve is conditional: If we charge the higher price, we would sell Q_1. If, instead, we charge a lower price (on all the units that we sell), we would sell Q_2.

So when we think about increasing the quantity sold by one unit, marginal revenue is affected in two ways. First, we sell one additional unit at the new market price. Second, all the previous units, which could have been sold at the higher price, now sell for less. Because of the lower price on all units sold, the marginal revenue of selling a unit is less than the price of that unit—and the marginal revenue curve is below the demand curve.

Tip: For a straight-line demand curve, MR and demand have the same vertical intercept. As output increases, marginal revenue decreases twice as fast as demand, so that the horizontal intercept of MR is halfway to the horizontal intercept of demand. You can see this in the [link].

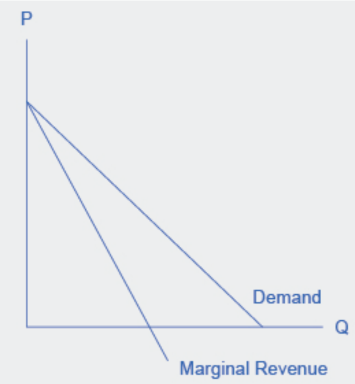


Figure 6: Because the market demand curve is conditional, the marginal revenue curve for a monopolist lies beneath the demand curve.

The Inefficiency of Monopoly

Most people criticize monopolies because they charge too high a price, but what economists object to is that monopolies do not supply enough output to be allocatively efficient. To understand why a monopoly is inefficient, it is useful to compare it with the benchmark model of perfect competition.

Allocative efficiency is a social concept. It refers to producing the optimal quantity of some output, the quantity where the marginal benefit to society of one more unit just equals the marginal cost. The rule of profit maximization in a world of perfect competition was for each firm to produce the quantity of output where P = MC, where the price (P) is a measure of how much buyers value the good and the marginal cost (MC) is a measure of what marginal units cost society to produce. Following this rule assures allocative efficiency. If P > MC, then the marginal benefit to society (as measured by P) is greater than the marginal cost to society of producing additional units, and a greater quantity should be produced. But in the case of monopoly, price is always greater than marginal cost at the profit-maximizing level of output, as can be seen by looking back at Figure 5. Thus, consumers will suffer from a monopoly because a lower quantity will be sold in the market, at a higher price, than would have been the case in a perfectly competitive market.

The problem of inefficiency for monopolies often runs even deeper than these issues, and also involves incentives for efficiency over longer periods of time. There are counterbalancing incentives here. On one side, firms may strive for new inventions and new intellectual property because they want to become monopolies and earn high profits—at least for a few years until the competition catches up. In this way, monopolies may come to exist because of competitive pressures on firms. However, once a barrier to entry is in place, a monopoly that does not need to fear competition can just produce the same old products in the same old way—while still ringing up a healthy rate of profit. John Hicks, who won the Nobel Prize for economics in 1972, wrote in 1935: “The best of all monopoly profits is a quiet life.” (Hicks: 1935: 8) He did not mean the comment in a complimentary way. He meant that monopolies may bank their profits and slack off on trying to please their customers.

When Telkom was the sole provider of telecommunications services up until early 2004,the payment plans and types of phones did not change much. However, in 2004 the Department of Communications redrafted the Electronics Communications Act which served to consolidate and redefine the landscape of telecommunications licensing in South Africa (both mobile and fixed). The Independent Communications Authority (ICASA) in 2009 licensed more than 400 independent operators with the Electronic Communications Network License (with the ability to self-provision) as well as issuing Electronic Communications Service Licenses for service deployment over infrastructure in the retail domain. Consequently Telkom is no longer strictly a monopoly

operator in South Africa, and faces competition from the second Fixed Network Operator Licensee, Neotel, as well as the three mobile operators, Vodacom, MTN and Cell-C. (Shevel and Prinsloo: 2013)

The freeing up of the South African telecommunications market resulted in services like call waiting, caller ID, three-way calling, voice mail through the phone company, mobile phones, and wireless connections to the Internet all becoming available. A wide range of payment plans was offered, as well. It was no longer true that all phones were black; instead, phones came in a wide variety of shapes and colors. The end of the telephone monopoly brought lower prices, a greater quantity of services, and also a wave of innovation aimed at attracting and pleasing customers.

Key Concepts and Summary

A monopolist is not a price taker, because when it decides what quantity to produce, it also determines the market price. For a monopolist, total revenue is relatively low at low quantities of output, because not much is being sold. Total revenue is also relatively low at very high quantities of output, because a very high quantity will sell only at a low price. Thus, total revenue for a monopolist will start low, rise, and then decline. The marginal revenue for a monopolist from selling additional units will decline. Each additional unit sold by a monopolist will push down the overall market price, and as more units are sold, this lower price applies to more and more units.

The monopolist will select the profit-maximizing level of output where MR = MC, and then charge the price for that quantity of output as determined by the market demand curve. If that price is above average cost, the monopolist earns positive profits.

Monopolists are not productively efficient, because they do not produce at the minimum of the average cost curve. Monopolists are not allocatively efficient, because they do not produce at the quantity where P = MC. As a result, monopolists produce less, at a higher average cost, and charge a higher price than would a combination of firms in a perfectly competitive industry. Monopolists also may lack incentives for innovation, because they need not fear entry.

Self-Check Questions

Exercise:

Problem:

Suppose demand for a monopoly's product falls so that its profit-maximizing price is below average variable cost. How much output should the firm supply? *Hint*: Draw the graph.

Solution:

If price falls below AVC, the firm will not be able to earn enough revenues even to cover its variable costs. In such a case, it will suffer a smaller loss if it shuts down and produces no output. By contrast, if it stayed in operation and produced the level of output where MR = MC, it would lose all of its fixed costs plus some variable costs. If it shuts down, it only loses its fixed costs.

Exercise:

Problem:

Imagine a monopolist could charge a different price to every customer based on how much he or she were willing to pay. How would this affect monopoly profits?

Solution:

This scenario is called “perfect price discrimination.” The result would be that the monopolist would produce more output, the same amount in fact as would be produced by a perfectly competitive industry. However, there would be no consumer surplus since each buyer is paying exactly what they think the product is worth. Therefore, the monopolist would be earning the maximum possible profits.

Review Questions

Exercise:

Problem:

How is the demand curve perceived by a perfectly competitive firm different from the demand curve perceived by a monopolist?

Exercise:

Problem:

How does the demand curve perceived by a monopolist compare with the market demand curve?

Exercise:

Problem: Is a monopolist a price taker? Explain briefly.

Exercise:

Problem: What is the usual shape of a total revenue curve for a monopolist? Why?

Exercise:

Problem: What is the usual shape of a marginal revenue curve for a monopolist? Why?

Exercise:

Problem:

How can a monopolist identify the profit-maximizing level of output if it knows its total revenue and total cost curves?

Exercise:

Problem:

How can a monopolist identify the profit-maximizing level of output if it knows its marginal revenue and marginal costs?

Exercise:

Problem:

When a monopolist identifies its profit-maximizing quantity of output, how does it decide what price to charge?

Exercise:

Problem: Is a monopolist allocatively efficient? Why or why not?

Exercise:

Problem:

How does the quantity produced and price charged by a monopolist compare to that of a perfectly competitive firm?

Critical Thinking Questions

Exercise:

Problem:

Imagine that you are managing a small firm and thinking about entering the market of a monopolist. The monopolist is currently charging a high price, and you have calculated that you can make a nice profit charging 10% less than the monopolist. Before you go ahead and challenge the monopolist, what possibility should you consider for how the monopolist might react?

Exercise:

Problem:

If a monopoly firm is earning profits, how much would you expect these profits to be diminished by entry in the long run?

Problems

Exercise:

Problem:

Draw the demand curve, marginal revenue, and marginal cost curves from [link], and identify the quantity of output the monopoly wishes to supply and the price it will charge. Suppose demand for the monopoly's product increases dramatically. Draw the new demand curve. What happens to the marginal revenue as a result of the increase in demand? What happens to the marginal cost curve? Identify the new profit-maximizing quantity and price. Does the answer make sense to you?

Exercise:

Problem:

Draw a monopolist's demand curve, marginal revenue, and marginal cost curves. Identify the monopolist's profit-maximizing output level. Now, think about a slightly higher level of output (say $Q_0 + 1$). According to the graph, is there any consumer willing to pay more than the marginal cost of that new level of output? If so, what does this mean?

References

Competition Tribunal South Africa. 2014. Outcome of the SASOL case. Outcome of competition tribunal case 5 June 2014. Available http://www.comptrib.co.za/publications/press-releases/outcome-of-the-sasol-case/ (Accessed 16 May 2016)

Hicks, J.R. 1935. Annual survey of economic theory: the theory of monopoly. Econometrica, 3(1): 8.

Shevel, A. and Prinsloo, L. 2013. Challenge for Telkom monopoly. Business day live, 6 October. Available http://www.bdlive.co.za/business/technology/2013/10/06/challenge-for-telkom-monopoly (Accessed: 16 May 2016)

Glossary

allocative efficiency
: producing the optimal quantity of some output; the quantity where the marginal benefit to society of one more unit just equals the marginal cost

marginal profit
: profit of one more unit of output, computed as marginal revenue minus marginal cost

Introduction to Monopolistic Competition and Oligopoly
class="introduction"
Competing Brands?

The laundry detergent market is one that is characterized neither as perfect competition nor monopoly. (Credit: modification of work by Pixel Drip/Flickr Creative Commons)

Note:

The Temptation to Defy the Law

Laundry detergent and bags of ice—products of industries that seem pretty unexciting, maybe even boring. Hardly! Both have been the center of secret meetings and secret deals worthy of a spy novel. In France, between 1997 and 2004, the top four laundry detergent producers (Proctor & Gamble, Henkel, Unilever, and Colgate-Palmolive) controlled about 90 percent of the French soap market. Officials from the soap firms were meeting secretly, in out-of-the-way, small cafés around Paris. Their goals: Stamp out competition and set prices.

In 2007 Tiger Brands, a company listed on the Johannesburg Stock Exchange (JSE), was fined R98,8 million by South Africa's competition authority after admitting that it made secret deals with rival companies to fix the price of bread (Serla: 2007). The complaint was made by independent bread distributors in the Western Cape province who complained to the commission in December 2006 that bakeries owned by Tiger Brands, the maker of Albany bread, Pioneer Foods and Premier

Foods had raised prices by between 30c and 35c a loaf a week before Christmas. The company was granted leniency against prosecution by agreeing to assist the Competition Commission with investigations into matters such as possible secret deals among grain millers.

If competing firms - by agreeing to act together - could meet their goals, it would enable each to act as though they were a single firm—in essence, a monopoly—and enjoy monopoly-size profits. The problem? In many parts of the world, including South Africa, it is illegal for firms to divide up markets and set prices collaboratively.

These two cases provide examples of markets that are characterized neither as perfect competition nor monopoly. Instead, these firms are competing in market structures that lie between the extremes of monopoly and perfect competition. How do they behave? Why do they exist? We will revisit this case later, to find out what happened.

Note:

Introduction to Monopolistic Competition and Oligopoly

In this chapter, you will learn about:

- Monopolistic Competition
- Oligopoly

Perfect competition and monopoly are at opposite ends of the competition spectrum. A perfectly competitive market has many firms selling identical products, who all act as price takers in the face of the competition. If you recall, price takers are firms that have no market power. They simply have to take the market price as given.

Monopoly arises when a single firm sells a product for which there are no close substitutes. Microsoft, for instance, has been considered a monopoly because of its domination of the operating systems market.

What about the vast majority of real world firms and organizations that fall between these extremes, firms that could be described as **imperfectly competitive**? What determines their behavior? They have more influence over the price they charge than perfectly competitive firms, but not as much as a monopoly would. What will they do?

One type of imperfectly competitive market is called **monopolistic competition**. Monopolistically competitive markets feature a large number of competing firms, but the products that they sell are not identical. Consider, as an example, the Mall of Africa in Midrand which is South Africa's biggest ever mall to have been built in a single phase (Steyn: 2013) . It boasts over 300 stores of which a significant number are clothing outlets which include international brands such as Cotton On, Forever 21, Forever New, River Island, Mango, Tommy Hilfiger and Versace. A number of local brands – Foschini Group, Mr Price and Truworths – have also chosen to make the Mall of Africa their new flagship stores, while anchors for the mall will include national favourites Woolworths, Edgars, Checkers and Game. Most of the markets that consumers encounter at the retail level are monopolistically competitive.

The other type of imperfectly competitive market is **oligopoly**. Oligopolistic markets are those dominated by a small number of firms. Commercial aircraft provides a good example: Boeing and Airbus each produce slightly less than 50% of the large commercial aircraft in the world. Another example is the cellular phone network market in South Africa: just a few firms such as Vodacom, MTN and Cell C dominate this industry. Oligopolies are characterized by high barriers to entry with firms choosing output, pricing, and other decisions strategically based on the decisions of the other firms in the market. In this chapter, we first explore how monopolistically competitive firms will choose their profit-maximizing level of output. We will then discuss oligopolistic firms, which face two conflicting temptations: to collaborate as if they were a single monopoly, or to individually compete to gain profits by expanding output levels and cutting prices. Oligopolistic markets and firms can also take on elements of monopoly and of perfect competition.

References

Tiger Brands slapped with R98m cartel fine. 2007. Mail and Guardian online, Business 12 Nov. Available http://mg.co.za/article/2007-11-12-tiger-brands-slapped-with-r98m-cartel-fine (Accessed 20 May 2016)

Steyn, L. 2013. “Mall of Africa: A New Shopping Experience.” Mail and Guardian, July 13, 2015, sec. Africa. Avaliable http://mg.co.za/article/2015-07-13-mall-of-africa-will-cement-sas-mall-superiority/ (Accessed 10 May 2016)

Monopolistic Competition

By the end of this section, you will be able to:

- Explain the significance of differentiated products
- Describe how a monopolistic competitor chooses price and quantity
- Discuss entry, exit, and efficiency as they pertain to monopolistic competition
- Analyze how advertising can impact monopolistic competition

Monopolistic competition involves many firms competing against each other, but selling products that are different in some way. Examples include stores that sell different styles of clothing; restaurants or grocery stores that sell different kinds of food; and even products like golf balls or beer that may be at least somewhat similar but differ in public perception because of advertising and brand names. There are over 8,000 restaurants and 1,200 hotels in South Africa (Mbendi Information Services, 2016). When products are distinctive, each firm has a mini-monopoly on its particular style or flavor or brand name. However, firms producing such products must also compete with other styles and flavors and brand names. The term "monopolistic competition" captures this mixture of mini-monopoly and tough competition, and the following Clear It Up feature introduces its derivation.

Note:
Who invented the theory of imperfect competition?
The theory of imperfect competition was developed by two economists independently but simultaneously in 1933. The first was Edward Chamberlin of Harvard University who published *The Economics of Monopolistic Competition*. The second was Joan Robinson of Cambridge University who published *The Economics of Imperfect Competition*. Robinson subsequently became interested in macroeconomics where she became a prominent Keynesian, and later a post-Keynesian economist.

Differentiated Products

A firm can try to make its products different from those of its competitors in several ways: physical aspects of the product, location from which the product is sold, marketing and packaging of the product, and perceptions of the product. Products that are distinctive in one of these ways are called **differentiated products**.

Physical aspects of a product include all the phrases you hear in advertisements: unbreakable bottle, non-stick surface, freezer-to-microwave, non-shrink, extra spicy, newly redesigned for your comfort. The location of a firm can also create a difference between producers in the minds of consumers. For example, a petrol station located at a busy intersection with heavy traffic can probably sell more fuel, because more cars drive by that corner. A parts supplier for a motor vehicle manufacturer may find it to be an advantage to be located close to the car factory.

Other aspects can differentiate a product too. These could include promises like a guarantee of satisfaction or money back, a reputation for high quality, services like free delivery, or offering a loan to purchase the product. Finally, product differentiation may occur in the minds of buyers. For example, many people could not tell the difference in taste between common varieties of beer or cigarettes if they were blindfolded but, because of past habits and advertising, they have strong preferences for certain brands. Advertising can play a role in shaping these consumer perceptions and hence their preferences.

The concept of differentiated products is closely related to the degree of variety that is available. If everyone in the economy wore only blue jeans, ate only white bread, and drank only tap water, then the markets for clothing, food, and drink would be much closer to perfectly competitive. The variety of styles, flavors, locations, and characteristics creates product differentiation and monopolistic competition.

Perceived Demand for a Monopolistic Competitor

A monopolistically competitive firm perceives a demand for its goods that is an intermediate case between monopoly and competition. Figure 1 offers a reminder that the demand curve as faced by a perfectly competitive firm is perfectly elastic or flat, because the perfectly competitive firm can sell any quantity it wishes at the prevailing market price. In contrast, the demand curve, as faced by a monopolist, is the market demand curve, since a monopolist is the only firm in the market, and hence is downward sloping.

Perceived Demand for Firms in Different Competitive Settings

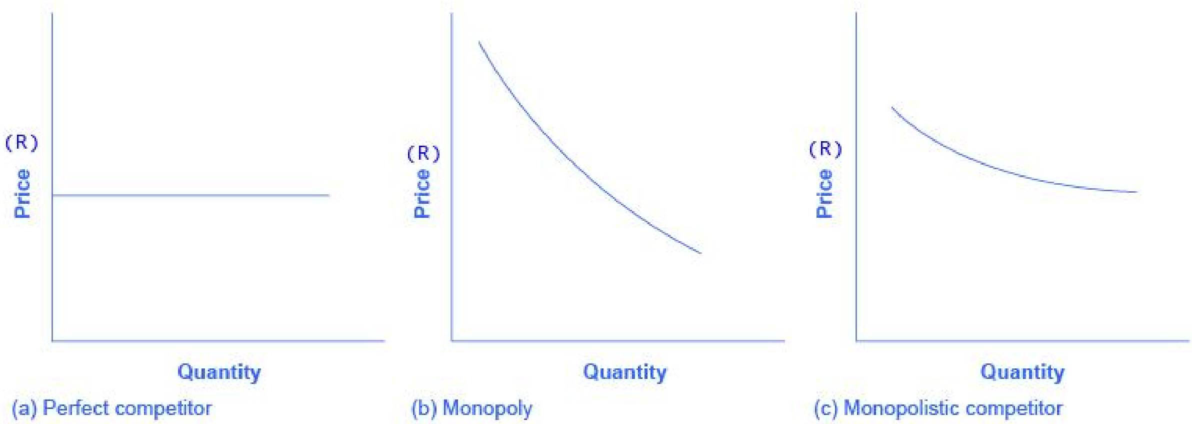


Figure 1: The demand curve faced by a perfectly competitive firm is perfectly elastic, meaning it can sell all the output it wishes at the prevailing market price. The demand curve faced by a monopoly is the market demand. It can sell more output only by decreasing the price it charges. The demand curve faced by a monopolistically competitive firm falls in between.

The demand curve as faced by a monopolistic competitor is not flat, but rather downward-sloping, which means that the monopolistic competitor can raise its price without losing all of its customers or lower the price and gain more customers. Since there are substitutes, the demand curve facing a monopolistically competitive firm is more elastic than that of a monopoly where there are no close substitutes. If a monopolist raises its price, some consumers will choose not to purchase its product—but they will then need to buy a completely different product. However, when a monopolistic competitor raises its price, some consumers will choose not to purchase the

product at all, but others will choose to buy a similar product from another firm. If a monopolistic competitor raises its price, it will not lose as many customers as would a perfectly competitive firm, but it will lose more customers than would a monopoly that raised its prices.

At a glance, the demand curves faced by a monopoly and by a monopolistic competitor look similar—that is, they both slope down. But the underlying economic meaning of these perceived demand curves is different, because a monopolist faces the market demand curve and a monopolistic competitor does not. Rather, a monopolistically competitive firm's demand curve is but one of many firms that make up the market demand curve.

How a Monopolistic Competitor Chooses Price and Quantity

The monopolistically competitive firm decides on its profit-maximizing quantity and price in much the same way as a monopolist. A monopolistic competitor, like a monopolist, faces a downward-sloping demand curve, and so it will choose some combination of price and quantity along its perceived demand curve.

As an example of a profit-maximizing monopolistic competitor, consider the Authentic Chinese Pizza store, which serves pizza slices with cheese, sweet and sour sauce, and your choice of vegetables and meats. Although Authentic Chinese Pizza must compete against other pizza businesses and restaurants, it has a differentiated product. The firm's perceived demand curve is downward sloping, as shown in Figure 2 and the first two columns of Table 1.

How a Monopolistic Competitor Chooses its Profit Maximizing Output and Price

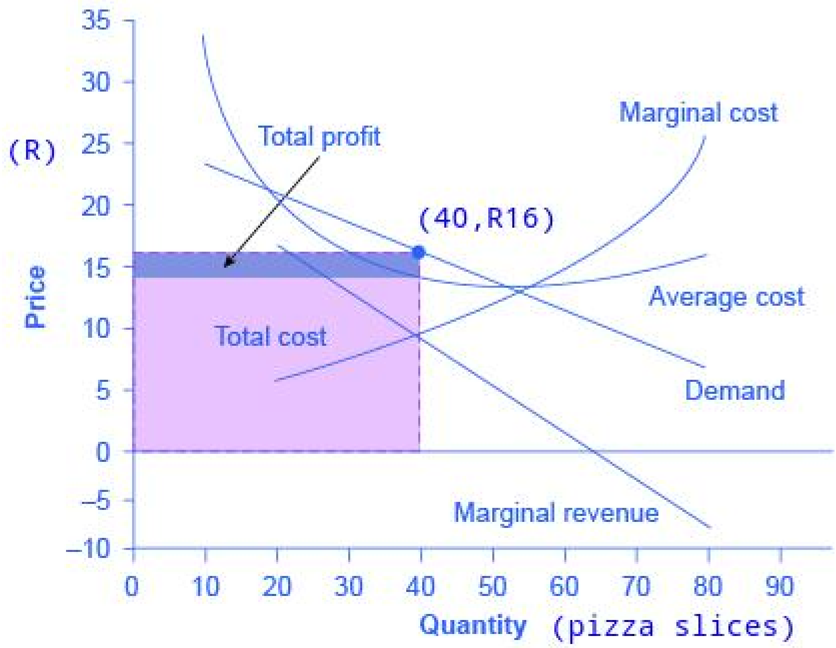


Figure 2: To maximize profits, the Authentic Chinese Pizza shop would choose a quantity where marginal revenue equals marginal cost, or Q where MR = MC. Here it would choose a quantity of 40 and a price of R16.

Quantity	Price (Rands)	Total Revenue (Rands)	Marginal Revenue (Rands)	Total Cost (Rands)	Marginal Cost (Rands)	Average Cost (Rands)
10	23	230	-	340	-	34
20	20	400	17	400	6	20
30	18	540	14	480	8	16
40	16	640	10	580	10	14.5
50	14	700	6	700	12	14
60	12	720	2	840	14	14
70	10	700	–2	1020	18	14.57
80	8	640	–6	1280	26	16

Table 1: Revenue and Cost Schedule

The combinations of price and quantity at each point on the demand curve can be multiplied to calculate the total revenue that the firm would receive,

which is shown in the third column of Table 1. The fourth column, marginal revenue, is calculated as the change in total revenue divided by the change in quantity. The final columns of Table 1 show total cost, marginal cost, and average cost. As always, marginal cost is calculated by dividing the change in total cost by the change in quantity, while average cost is calculated by dividing total cost by quantity. The following Work It Out feature shows how these firms calculate how much of its product to supply at what price.

Note:

How a Monopolistic Competitor Determines How Much to Produce and at What Price

The process by which a monopolistic competitor chooses its profit-maximizing quantity and price resembles closely how a monopoly makes these decisions process. First, the firm selects the profit-maximizing quantity to produce. Then the firm decides what price to charge for that quantity.

Step 1. The monopolistic competitor determines its profit-maximizing level of output. In this case, the Authentic Chinese Pizza company will determine the profit-maximizing quantity to produce by considering its marginal revenues and marginal costs. Two scenarios are possible:

- If the firm is producing at a quantity of output where marginal revenue exceeds marginal cost, then the firm should keep expanding production, because each marginal unit is adding to profit by bringing in more revenue than its cost. In this way, the firm will produce up to the quantity where MR = MC.
- If the firm is producing at a quantity where marginal costs exceed marginal revenue, then each marginal unit is costing more than the revenue it brings in, and the firm will increase its profits by reducing the quantity of output until MR = MC.

In this example, MR and MC intersect at a quantity of 40, which is the profit-maximizing level of output for the firm.

Step 2. The monopolistic competitor decides what price to charge. When the firm has determined its profit-maximizing quantity of output, it can then look to its perceived demand curve to find out what it can charge for

that quantity of output. On the graph, this process can be shown as a vertical line reaching up through the profit-maximizing quantity until it hits the firm's perceived demand curve. For Authentic Chinese Pizza, it should charge a price of R16 per pizza slice for a quantity of 40 slices.
Once the firm has chosen price and quantity, it's in a position to calculate total revenue, total cost, and profit. At a quantity of 40, the price of R16 per slice lies above the average cost curve, so the firm is making economic profits. From Table 1 we can see that, at an output of 40 slices, the firm's total revenue is R640 and its total cost is R580, so total profits are R60. In Figure 2, the firm's total revenues are the rectangle with the quantity of 40 on the horizontal axis and the price of R16 on the vertical axis. The firm's total costs are the light shaded rectangle with the same quantity of 40 on the horizontal axis but the average cost of R14.50 per slice on the vertical axis. Total profits are total revenues minus total costs, which is the shaded area above the average cost curve.

Although the process by which a monopolistic competitor makes decisions about quantity and price is similar to the way in which a monopolist makes such decisions, two differences are worth remembering. First, although both a monopolist and a monopolistic competitor face downward-sloping demand curves, the monopolist's perceived demand curve is the market demand curve, while the perceived demand curve for a monopolistic competitor is based on the extent of its product differentiation and how many competitors it faces. Second, a monopolist is surrounded by barriers to entry and need not fear entry, but a monopolistic competitor who earns profits must expect the entry of firms with similar, but differentiated, products.

Monopolistic Competitors and Entry

If one monopolistic competitor earns positive economic profits, other firms will be tempted to enter the market. A petrol station with a great location must worry that other fuel stations might open across the street or down the road—and perhaps the new fuel stations will sell coffee or have a car-wash or some other attraction to attract customers. A successful restaurant with a

unique braai sauce must be concerned that other restaurants will try to copy the sauce or offer their own unique recipes. A laundry detergent with a great reputation for quality must be concerned that other competitors may seek to build their own reputations.

The entry of other firms into the same general market (like fuel, restaurants, or detergent) shifts the demand curve faced by a monopolistically competitive firm. As more firms enter the market, the quantity demanded at a given price for any particular firm will decline, and the firm's perceived demand curve will shift to the left. As a firm's perceived demand curve shifts to the left, its marginal revenue curve will shift to the left, too. The shift in marginal revenue will change the profit-maximizing quantity that the firm chooses to produce, since marginal revenue will then equal marginal cost at a lower quantity.

Figure 3 (a) shows a situation in which a monopolistic competitor was earning a profit with its original perceived demand curve (D_0). The intersection of the marginal revenue curve (MR_0) and marginal cost curve (MC) occurs at point S, corresponding to quantity Q_0, which is associated on the demand curve at point T with price P_0. The combination of price P_0 and quantity Q_0 lies above the average cost curve, which shows that the firm is earning positive economic profits.

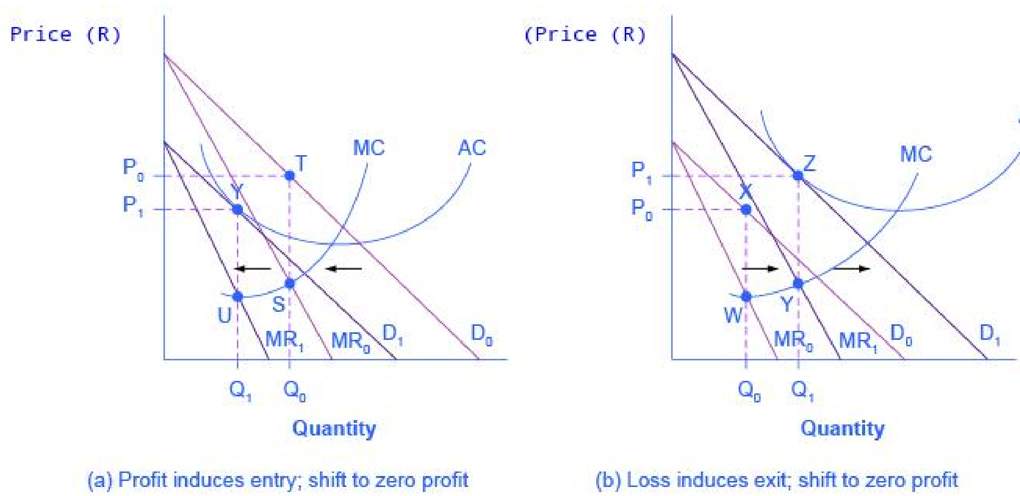


Figure 3: (a) At P0 and Q0, the monopolistically competitive firm shown in this figure is making a positive economic profit. This is clear because if you follow the dotted line above Q0, you can see that price is above average cost. Positive economic profits attract competing firms to the industry, driving the original firm's demand down to D1. At the new equilibrium quantity (P1, Q1), the original firm is earning zero economic profits, and entry into the industry ceases. In (b) the opposite occurs. At P0 and Q0, the firm is losing money. If you follow the dotted line above Q0, you can see that average cost is above price. Losses induce firms to leave the industry. When they do, demand for the original firm rises to D1, where once again the firm is earning zero economic profit.

Unlike a monopoly, with its high barriers to entry, a monopolistically competitive firm with positive economic profits will attract competition. When another competitor enters the market, the original firm's perceived demand curve shifts to the left, from D_0 to D_1, and the associated marginal revenue curve shifts from MR_0 to MR_1. The new profit-maximizing output is Q_1, because the intersection of the MR_1 and MC now occurs at point U. Moving vertically up from that quantity on the new demand curve, the optimal price is at P_1.

As long as the firm is earning positive economic profits, new competitors will continue to enter the market, reducing the original firm's demand and marginal revenue curves. The long-run equilibrium is shown in the figure at point Y, where the firm's perceived demand curve touches the average cost curve. When price is equal to average cost, economic profits are zero. Thus, although a monopolistically competitive firm may earn positive economic profits in the short term, the process of new entry will drive down economic profits to zero in the long run. Remember that zero economic profit is not equivalent to zero accounting profit. A zero economic profit means the

firm’s accounting profit is equal to what its resources could earn in their next best use. Figure 3 (b) shows the reverse situation, where a monopolistically competitive firm is originally losing money. The adjustment to long-run equilibrium is similar to the previous example. The economic losses lead to firms exiting, which will result in increased demand for this particular firm, and consequently lower losses. Firms exit up to the point where there are no more losses in this market, for example when the demand curve touches the average cost curve, as in point Z.

Monopolistic competitors can make an economic profit or loss in the short run, but in the long run, entry and exit will drive these firms toward a zero economic profit outcome. However, the zero economic profit outcome in monopolistic competition looks different from the zero economic profit outcome in perfect competition in several ways relating both to efficiency and to variety in the market.

Monopolistic Competition and Efficiency

The long-term result of entry and exit in a perfectly competitive market is that all firms end up selling at the price level determined by the lowest point on the average cost curve. This outcome is why perfect competition displays productive efficiency: goods are being produced at the lowest possible average cost. However, in monopolistic competition, the end result of entry and exit is that firms end up with a price that lies on the downward-sloping portion of the average cost curve, not at the very bottom of the AC curve. Thus, monopolistic competition will not be productively efficient.

In a perfectly competitive market, each firm produces at a quantity where price is set equal to marginal cost, both in the short run and in the long run. This outcome is why perfect competition displays allocative efficiency: the social benefits of additional production, as measured by the marginal benefit, which is the same as the price, equal the marginal costs to society of that production. In a monopolistically competitive market, the rule for maximizing profit is to set MR = MC—and price is higher than marginal revenue, not equal to it because the demand curve is downward sloping. When P > MC, which is the outcome in a monopolistically competitive market, the benefits to society of providing additional quantity, as measured

by the price that people are willing to pay, exceed the marginal costs to society of producing those units. A monopolistically competitive firm does not produce more, which means that society loses the net benefit of those extra units. This is the same argument we made about monopoly, but in this case to a lesser degree. Thus, a monopolistically competitive industry will produce a lower quantity of a good and charge a higher price for it than would a perfectly competitive industry. See the following Clear It Up feature for more detail on the impact of demand shifts.

Note:

Why does a shift in perceived demand cause a shift in marginal revenue?

The combinations of price and quantity at each point on a firm's perceived demand curve are used to calculate total revenue for each combination of price and quantity. This information on total revenue is then used to calculate marginal revenue, which is the change in total revenue divided by the change in quantity. A change in perceived demand will change total revenue at every quantity of output and in turn, the change in total revenue will shift marginal revenue at each quantity of output. Thus, when entry occurs in a monopolistically competitive industry, the perceived demand curve for each firm will shift to the left, because a smaller quantity will be demanded at any given price. Another way of interpreting this shift in demand is to notice that, for each quantity sold, a lower price will be charged. Consequently, the marginal revenue will be lower for each quantity sold—and the marginal revenue curve will shift to the left as well. Conversely, exit causes the perceived demand curve for a monopolistically competitive firm to shift to the right and the corresponding marginal revenue curve to shift right, too.

A monopolistically competitive industry does not display productive and allocative efficiency in either the short run, when firms are making economic profits and losses, or in the long run, when firms are earning zero profits.

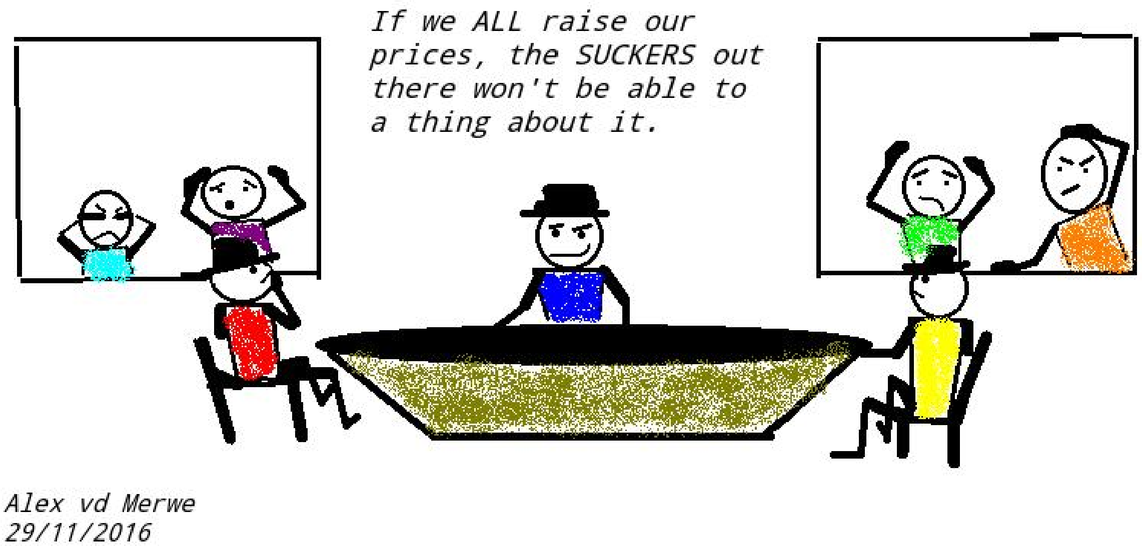


Figure 4. Oligopolists colluding in the boardroom.

The Benefits of Variety and Product Differentiation

Even though monopolistic competition does not provide productive efficiency or allocative efficiency, it does have benefits of its own. Product differentiation is based on variety and innovation. Many people would prefer to live in an economy with many kinds of clothes, foods, and car styles; not in a world of perfect competition where everyone will always wear blue jeans and white shirts, eat only spaghetti with plain red sauce, and drive an identical model of car. Many people would prefer to live in an economy where firms are struggling to figure out ways of attracting customers by methods like friendlier service, free delivery, guarantees of quality, variations on existing products, and a better shopping experience.

Economists have struggled, with only partial success, to address the question of whether a market-oriented economy produces the optimal amount of variety. Critics of market-oriented economies argue that society does not really need dozens of different athletic shoes or breakfast cereals or automobiles. They argue that much of the cost of creating such a high degree of product differentiation, and then of advertising and marketing this differentiation, is socially wasteful—that is, most people would be just as

happy with a smaller range of differentiated products produced and sold at a lower price. Defenders of a market-oriented economy respond that if people do not want to buy differentiated products or highly advertised brand names, no one is forcing them to do so. Moreover, they argue that consumers benefit substantially when firms seek short-term profits by providing differentiated products. This controversy may never be fully resolved, in part because deciding on the optimal amount of variety is very difficult, and in part because the two sides often place different values on what variety means for consumers. Read the following Clear It Up feature for a discussion on the role that advertising plays in monopolistic competition.

Note:

How does advertising impact monopolistic competition?

The United States economy spent about $180.12 billion on advertising in 2014, according to eMarketer.com. Roughly one third of this was television advertising, and another third was divided roughly equally between Internet, newspapers, and radio. The remaining third was divided up between direct mail, magazines, telephone directory yellow pages, and billboards. Mobile devices are increasing the opportunities for advertisers.

Advertising is all about explaining to people, or making people believe, that the products of one firm are differentiated from the products of another firm. In the framework of monopolistic competition, there are two ways to conceive of how advertising works: either advertising causes a firm’s perceived demand curve to become more inelastic (that is, it causes the perceived demand curve to become steeper); or advertising causes demand for the firm’s product to increase (that is, it causes the firm’s perceived demand curve to shift to the right). In either case, a successful advertising campaign may allow a firm to sell either a greater quantity or to charge a higher price, or both, and thus increase its profits.

However, economists and business owners have also long suspected that much of the advertising may only offset other advertising. Economist A. C. Pigou wrote the following back in 1920 in his book, *The Economics of Welfare*:

> "It may happen that expenditures on advertisement made by competing monopolists [that is, what we now call monopolistic competitors] will simply neutralise one another, and leave the industrial position exactly as it would have been if neither had expended anything. For, clearly, if each of two rivals makes equal efforts to attract the favour of the public away from the other, the total result is the same as it would have been if neither had made any effort at all."

So it is not only the actions of consumers (demand) that may affect market outcomes, but also the actions of rival firms. In the next section on oligopoly (what a word!), we see how firms collude (meet in secret) to manipulate market prices.

Key Concepts and Summary

Monopolistic competition refers to a market where many firms sell differentiated products. Differentiated products can arise from characteristics of the good or service, location from which the product is sold, intangible aspects of the product, and perceptions of the product.

The perceived demand curve for a monopolistically competitive firm is downward-sloping, which shows that it is a price maker and chooses a combination of price and quantity. However, the perceived demand curve for a monopolistic competitor is more elastic than the perceived demand curve for a monopolist, because the monopolistic competitor has direct competition, unlike the pure monopolist. A profit-maximizing monopolistic competitor will seek out the quantity where marginal revenue is equal to marginal cost. The monopolistic competitor will produce that level of output and charge the price that is indicated by the firm’s demand curve.

If the firms in a monopolistically competitive industry are earning economic profits, the industry will attract entry until profits are driven down to zero in the long run. If the firms in a monopolistically competitive industry are suffering economic losses, then the industry will experience exit of firms until economic profits are driven up to zero in the long run.

A monopolistically competitive firm is not productively efficient because it does not produce at the minimum of its average cost curve. A monopolistically competitive firm is not allocatively efficient because it does not produce where P = MC, but instead produces where P > MC. Thus, a monopolistically competitive firm will tend to produce a lower quantity at a higher cost and to charge a higher price than a perfectly competitive firm.

Monopolistically competitive industries do offer benefits to consumers in the form of greater variety and incentives for improved products and services. There is some controversy over whether a market-oriented economy generates too much variety and results in unnecessarily higher prices.

Self-Check Questions

Exercise:

Problem:

Suppose that, due to a successful advertising campaign, a monopolistic competitor experiences an increase in demand for its product. How will that affect the price it charges and the quantity it supplies?

Solution:

An increase in demand will manifest itself as a rightward shift in the demand curve, and a rightward shift in marginal revenue. The shift in marginal revenue will cause a movement up the marginal cost curve to the new intersection between MR and MC at a higher level of output. The new price can be read by drawing a line up from the new output level to the new demand curve, and then over to the vertical axis. The new price should be higher. The increase in quantity will cause a movement along the average cost curve to a possibly higher level of average cost. The price, though, will increase more, causing an increase in total profits.

Exercise:

Problem:

Continuing with the scenario outlined in question 1, in the long run, the positive economic profits earned by the monopolistic competitor will attract a response either from existing firms in the industry or firms outside. As those firms capture the original firm's profit, what will happen to the original firm's profit-maximizing price and output levels?

Solution:

As long as the original firm is earning positive economic profits, other firms will respond in ways that take away the original firm's profits. This will manifest itself as a decrease in demand for the original firm's product, a decrease in the firm's profit-maximizing price and a decrease in the firm's profit-maximizing level of output, essentially unwinding the process described in the answer to exercise 1. In the long-run equilibrium, all firms in monopolistically competitive markets will earn zero economic profits.

Review Questions

Exercise:

Problem:

What is the relationship between product differentiation and monopolistic competition?

Exercise:

Problem:

How is the perceived demand curve for a monopolistically competitive firm different from the perceived demand curve for a monopoly or a perfectly competitive firm?

Exercise:

Problem:

How does a monopolistic competitor choose its profit-maximizing quantity of output and price?

Exercise:

Problem:

How can a monopolistic competitor tell whether the price it is charging will cause the firm to earn profits or experience losses?

Exercise:

Problem:

If the firms in a monopolistically competitive market are earning economic profits or losses in the short run, would you expect them to continue doing so in the long run? Why?

Exercise:

Problem:

Is a monopolistically competitive firm productively efficient? Is it allocatively efficient? Why or why not?

Critical Thinking Questions

Exercise:

Problem:

Aside from advertising, how can monopolistically competitive firms increase demand for their products?

Exercise:

Problem:

Make a case for why monopolistically competitive industries never reach long-run equilibrium.

Exercise:

Problem:

Would you rather have efficiency or variety? That is, one opportunity cost of the variety of products we have is that each product costs more per unit than if there were only one kind of product of a given type, like shoes. Perhaps a better question is, “What is the right amount of variety? Can there be too many varieties of shoes, for example?”

References

Chamberlin, E.H. 1933. The economics of monopolistic competition. Cambridge: Harvard University Press.

Mbendi Information Services. 2016.“Travel, Tourism and Recreation in South Africa.” Available https://www.mbendi.com/indy/trvl/af/sa/p0005.htm.(Accessed 29 Nov 2016)

Pigou, A C. 1920. The economics of welfare. London: Macmillan.

Robinson, J. 1933. The economics of imperfect competition. New York: Macmillan.

Total US Ad Spending to See Largest Increase Since 2004: Mobile advertising leads growth; will surpass radio, magazines and newspapers this year. 2014. eMarketer July 2. Available http://www.emarketer.com/Article/Total-US-Ad-Spending-See-Largest-Increase-Since-2004/1010982 (Accessed 20 May 2016)

Glossary

differentiated product
: a product that is perceived by consumers as distinctive in some way

imperfectly competitive

firms and organizations that fall between the extremes of monopoly and perfect competition

monopolistic competition
many firms competing to sell similar but differentiated products

oligopoly
when a few large firms have all or most of the sales in an industry

Oligopoly

By the end of this section, you will be able to:

- Explain why and how oligopolies exist
- Contrast collusion and competition
- Interpret and analyze the prisoner's dilemma diagram
- Evaluate the tradeoffs of imperfect competition

Many purchases that individuals make at the retail level are produced in markets that are neither perfectly competitive, monopolies, nor monopolistically competitive. Rather, they are oligopolies. Oligopoly arises when a small number of large firms have all or most of the sales in an industry. There are many examples of oligopoly. In South Africa these include the motor vehicle industry, the banking sector, the cellular phone network market and the airline industry. Oligopolistic firms are like cats in a bag. They can either scratch each other to pieces or cuddle up and get comfortable with one another. If oligopolists compete hard, they may end up acting very much like perfect competitors, driving down costs and leading to zero profits for all. If oligopolists cooperate with each other, they may effectively act like a monopoly and succeed in pushing up prices and earning consistently high levels of profit. Oligopolies are typically characterized by mutual interdependence where various decisions such as output, price, advertising, and so on, depend on the decisions of the other firm(s). Analyzing the choices of oligopolistic firms about pricing and quantity produced involves considering the pros and cons of competition versus collusion at a given point in time.

Why Do Oligopolies Exist?

A combination of the barriers to entry that create monopolies and the product differentiation that characterizes monopolistic competition can create the setting for an oligopoly. For example, when a government grants a patent for an invention to one firm, it may create a monopoly. When the government grants patents to, for example, three different pharmaceutical companies that each has its own drug for reducing high blood pressure, those three firms may become an oligopoly.

Similarly, a natural monopoly will arise when the quantity demanded in a market is only large enough for a single firm to operate at the minimum of the long-run average cost curve. In such a setting, the market has room for only one firm, because no smaller firm can operate at a low enough average cost to compete, and no larger firm could sell what it produced given the quantity demanded in the market.

Quantity demanded in the market may also be two or three times the quantity needed to produce at the minimum of the average cost curve—which means that the market would have room for only two or three oligopoly firms (and they need not produce differentiated products). Again, smaller firms would have higher average costs and be unable to compete, while additional large firms would produce such a high quantity that they would not be able to sell it at a profitable price. This combination of economies of scale and market demand creates a major barrier to entry which led, for example, to the Boeing-Airbus oligopoly for large passenger aircraft.

The product differentiation at the heart of monopolistic competition can also play a role in creating oligopoly. For example, firms may need to reach a certain minimum size before they are able to spend enough on advertising and marketing to create a recognizable brand name. The problem in competing with, say, Coca-Cola or Pepsi is not that producing fizzy drinks is technologically difficult, but rather that creating a brand name and marketing effort to equal Coke or Pepsi is an enormous task. This is something that South African company Shoreline Beverages with its Coo-ee brand has achieved in managing to capture 15% of the carbonated soft drinks market in KwaZulu-Natal and an average of 35% of the market in the channels in which it operates (Inggs: 2009). Coo-ee is priced at about 15% less than ABI (Amalgamated Beverage Industries) brands such as Sparletta, Fanta and Coke.

Collusion or Competition?

When oligopoly firms in a certain market decide what quantity to produce and what price to charge, they face a temptation to act as if they were a monopoly. By acting together, oligopolistic firms can hold down industry

output, charge a higher price, and divide up the profit among themselves. When firms act together in this way to reduce output and keep prices high, it is called **collusion**. A group of firms that have a formal agreement to collude to produce the monopoly output and sell at the monopoly price is called a **cartel**.

The desire of businesses to avoid competing so that they can instead raise the prices that they charge and earn higher profits has been well understood by economists. Adam Smith wrote in *Wealth of Nations* in 1776: “People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices.” (Smith, A.: 1776: Vol.1, Ch.10, Para 82)

Even when oligopolists recognize that they would benefit as a group by acting like a monopoly, each individual oligopoly faces a private temptation to produce just a slightly higher quantity and earn slightly higher profit—while still counting on the other oligopolists to hold down their production and keep prices high. If at least some oligopolists give in to this temptation and start producing more, then the market price will fall. Indeed, a small handful of oligopoly firms may end up competing so fiercely that they all end up earning zero economic profits—as if they were perfect competitors.

The Prisoner’s Dilemma

Because of the complexity of oligopoly, which is the result of mutual interdependence among firms, there is no single, generally-accepted theory of how oligopolies behave, in the same way that we have theories for all the other market structures. Instead, economists use **game theory**, a branch of mathematics that analyzes situations in which players must make decisions and then receive payoffs based on what other players decide to do. Game theory has found widespread applications in the social sciences, as well as in business, law, and military strategy.

The **prisoner’s dilemma** is a scenario in which the gains from cooperation are larger than the rewards from pursuing self-interest. It applies well to oligopoly. The story behind the prisoner’s dilemma goes like this:

"Two criminals (who were acting together) are arrested. When they are taken to the police station, they refuse to say anything and are put in separate interrogation rooms. Eventually, a police officer enters the room where Prisoner A is being held and says: “You know what? Your partner in the other room is confessing. So your partner is going to get a light prison sentence of just one year, and because you’re remaining silent, the judge is going to stick you with eight years in prison. Why don’t you get smart? If you confess, too, we’ll cut your jail time down to five years, and your partner will get five years, also.” Over in the next room, another police officer is giving exactly the same speech to Prisoner B. What the police officers do not say is that if both prisoners remain silent, the evidence against them is not especially strong, and the prisoners will end up with only two years in jail each."

The game theory situation facing the two prisoners is shown in Table 1. To understand the dilemma, first consider the choices from Prisoner A’s point of view. If A believes that B will confess, then A ought to confess, too, so as to not get stuck with the eight years in prison. But if A believes that B will not confess, then A will be tempted to act selfishly and confess, so as to serve only one year. The key point is that A has an incentive to confess regardless of what choice B makes! B faces the same set of choices, and thus will have an incentive to confess regardless of what choice A makes. Confess is considered the dominant strategy or the strategy an individual (or firm) will pursue regardless of the other individual’s (or firm’s) decision. The result is that if prisoners pursue their own self-interest, both are likely to confess, and end up doing a total of 10 years of jail time between them.

	Prisoner B	
	Remain Silent (cooperate	Confess (do not cooperate

		with other prisoner)	with other prisoner)
Prisoner A	Remain Silent (cooperate with other prisoner)	A gets 2 years, B gets 2 years	A gets 8 years, B gets 1 year
	Confess (do not cooperate with other prisoner)	A gets 1 year, B gets 8 years	A gets 5 years B gets 5 years

The Prisoner’s Dilemma Problem

The game is called a dilemma because if the two prisoners had cooperated by both remaining silent, they would only have had to serve a total of four years of jail time between them. If the two prisoners can work out some way of cooperating so that neither one will confess, they will both be better off than if they each follow their own individual self-interest, which in this case leads straight into longer jail terms.

The Oligopoly Version of the Prisoner’s Dilemma

The members of an oligopoly can face a prisoner’s dilemma, also. If each of the oligopolists cooperates in holding down output, then high monopoly profits are possible. Each oligopolist, however, must worry that while it is holding down output, other firms are taking advantage of the high price by raising output and earning higher profits. Table 2 shows the prisoner’s dilemma for a two-firm oligopoly—known as a **duopoly**. If Firms A and B both agree to hold down output, they are acting together as a monopoly and will each earn R1,000 in profits. However, both firms’ dominant strategy is to increase output, in which case each will earn R400 in profits.

		Firm B	
		Hold Down Output (cooperate with other firm)	Increase Output (do not cooperate with other firm)
Firm A	Hold Down Output (cooperate with other firm)	A gets R1,000, B gets R1,000	A gets R200, B gets R1,500
	Increase Output (do not cooperate with other firm)	A gets R1,500, B gets R200	A gets R400, B gets R400

A Prisoner's Dilemma for Oligopolists

Can the two firms trust each other? Consider the situation of Firm A:

- If A thinks that B will cheat on their agreement and increase output, then A will increase output, too, because for A the profit of R400 when both firms increase output (the bottom right-hand choice in [link]) is better than a profit of only R200 if A keeps output low and B raises output (the upper right-hand choice in the table).
- If A thinks that B will cooperate by holding down output, then A may seize the opportunity to earn higher profits by raising output. After all, if B is going to hold down output, then A can earn R1,500 in profits by expanding output (the bottom left-hand choice in the table) compared with only R1,000 by holding down output as well (the upper left-hand choice in the table).

Thus, firm A will reason that it makes sense to expand output if B holds down output and that it also makes sense to expand output if B raises output. Again, B faces a parallel set of decisions.

The result of this prisoner’s dilemma is often that even though A and B could make the highest combined profits by cooperating in producing a lower level of output and acting like a monopolist, the two firms may well end up in a situation where they each increase output and earn only R400 each in profits. The following Clear It Up feature discusses one cartel scandal in particular.

Note:
South Africa's Construction Cartel
Long-running and hugely profitable organized co-operation between major construction companies in South Africa started to fall apart because of a breakdown of arrangements between participants when it came to paying "loser fees" (Benjamin and De Wet: 2013) connected with the Coega Harbour project and also with respect to the construction of Soccer World Cup Stadiums. In one incident Concor,a large construction firm withheld "loser fees" from competitors that had helped it to win a major contract.The agreement (collusion) was that Concor's competitors would submit tenders for projects that they had agreed (for a fee) would lose against Concor's bid. However, after Concor had won the contracts, it refused to pay its competitors the "loser" fees.
Trust within the cartel was broken with the consequence that Concor's competitors agreed to testify against it in hearings by the Competition Commission. Their affidavits put before prosecutors showed that losing bidders, which intentionally priced themselves out of contracts but nonetheless competed to provide a semblance of reality to the tender process, were regularly paid the cost of submitting "fake" tenders. The cartel companies faced a total penalty bill of R1.5 billion for bid rigging.

How to Enforce Cooperation

How can parties who find themselves in a "prisoner’s dilemma" situation avoid the undesired outcome and cooperate with each other? The way out of a "prisoner’s dilemma" is to find a way to penalize those who do not cooperate.

Perhaps the easiest approach for colluding oligopolists, as you might imagine, would be to sign a contract with each other that they will hold output low and keep prices high. If a group of South African companies signed such a contract, however, it would be illegal. Certain international organizations, like the nations that are members of the Organization of Petroleum Exporting Countries (OPEC), have signed international agreements to act like a monopoly, hold down output, and keep prices high so that all of the countries can make high profits from oil exports. Such agreements, however, because they fall in a gray area of international law, are not legally enforceable. If Nigeria, for example, decides to start cutting prices and selling more oil, Saudi Arabia cannot sue Nigeria in court and force it to stop.

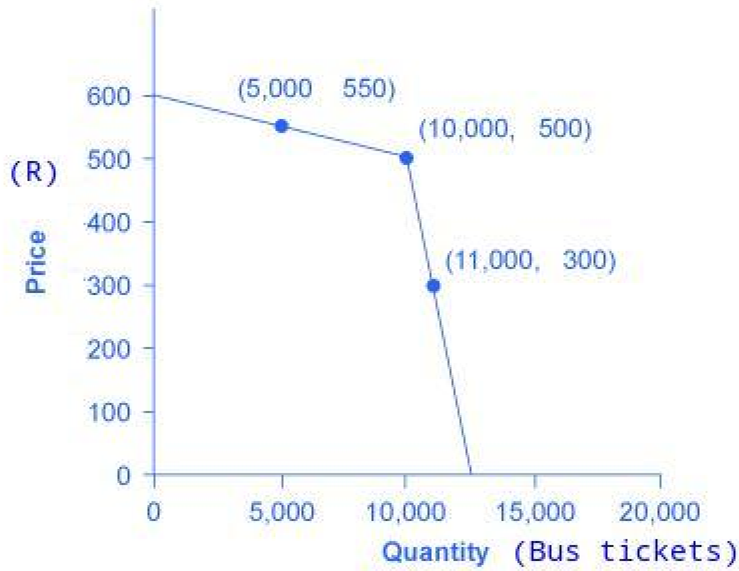
Note:
Visit the Organization of Petroleum Exporting Countries website and learn more about its history and how it defines itself.

Because oligopolists cannot sign a legally enforceable contract to act like a monopoly, the firms may instead keep close watch on what other firms are producing and charging. Alternatively, oligopolists may choose to act in a way that generates pressure on each firm to stick to its agreed quantity of output.

One example of the pressure these firms can exert on one another is the **kinked or bent demand curve**, in which competing oligopoly firms commit to match price cuts, but not price increases. This situation is shown

in Figure 1. Say that a KwaZulu-Natal bus transport company has agreed with the rest of a cartel (so they operate as an oligopoly) to provide a quantity of 10,000 seats on the Durban to Johannesburg route, at a price of R500 for a one-way ticket. This choice defines the kink in the firm's perceived demand curve. The reason that the firm faces a kink in its demand curve is because of how the other firms react to changes in the firm's price. If the firm decides to sell more tickets and cut its price, the other members of the cartel will immediately match any price cuts—and therefore, a lower price brings very little increase in quantity sold.

If one firm cuts its price to R300 per ticket, it will be able to sell only 11,000 seats. However, if the firm seeks to raise prices, the other firms in the cartel will not raise their prices, and so the firm that raised prices will lose a considerable share of sales. For example, if the firm raises its price to R550 per ticket, its sales drop to 5,000 seats sold. Thus, if firms always match price cuts by other firms in the oligopoly/cartel, but do not match price increases, then none of the firms in the oligopoly/cartel will have a strong incentive to change prices, since the potential gains are minimal. This strategy can work like a silent form of cooperation, in which the cartel successfully manages to hold down output, increase price, and share a monopoly level of profits even without any legally enforceable agreement.
A Kinked Demand Curve



Consider a member firm in an oligopoly

cartel that is supposed to produce a quantity of 10,000 tickets and sell at a price of R500/ticket. The other members of the cartel can encourage this firm to honor its commitments by acting so that the firm faces a kinked demand curve. If the firm attempts to expand output and reduce price slightly, other firms also cut prices immediately—so if the firm expands output to 11,000, the price per unit falls dramatically, to R300. On the other side, if the oligopoly attempts to raise its price, other firms will not do so, so if the firm raises its price to R550, its sales decline sharply to 5,000 tickets. Thus, the members of a cartel can discipline each other to stick to the pre-agreed levels of quantity and price through a strategy of matching all price cuts but not matching any price increases.

Many real-world oligopolies, forced by economic changes, legal and political pressures, and the egos of their top executives, go through episodes of cooperation and competition. If oligopolies could sustain cooperation with each other on output and pricing, they could earn profits as if they were a single monopoly. However, each firm in an oligopoly has an incentive to produce more and grab a bigger share of the overall market; when firms start behaving in this way, the market outcome in terms of prices and quantity can be similar to that of a highly competitive market.

Tradeoffs of Imperfect Competition

Monopolistic competition is probably the single most common market structure in most economies including South Africa's. It provides powerful incentives for innovation, as firms seek to earn profits in the short run,

while entry assures that firms do not earn economic profits in the long run. However, monopolistically competitive firms do not produce at the lowest point on their average cost curves. In addition, the endless search to impress consumers through product differentiation may lead to excessive social expenses on advertising and marketing with consequent negative effects on the environment (pollution arising from unnecessary packaging).

Oligopoly is probably the second most common market structure. When oligopolies result from patented innovations or from taking advantage of economies of scale to produce at low average cost, they may provide considerable benefit to consumers. Oligopolies are often protected by significant barriers to entry, which enable the oligopolists to earn sustained profits over long periods of time. Oligopolists also do not typically produce at the minimum of their average cost curves. When they lack vibrant competition, they may lack incentives to provide innovative products and high-quality service.

The task of government or public policy with regard to competition is to sort through these multiple realities, attempting to encourage behavior that is beneficial to the broader society and to discourage behavior that only adds to the profits of a few large companies, with no corresponding benefit to consumers. South Africa's Competition Commission and Competition Tribunal are charged with the responsibility of monitoring anti-competitive behavior by firms.

Key Concepts and Summary

An oligopoly is a situation where a few firms sell most or all of the goods in a market. Oligopolists earn their highest profits if they can band together as a cartel and act like a monopolist by reducing output and raising price. Since each member of the oligopoly can benefit individually from expanding output, such collusion often breaks down—especially since explicit collusion is illegal.

The "prisoner’s dilemma" is an example of game theory. It shows how, in certain situations, all sides can benefit from cooperative behavior rather

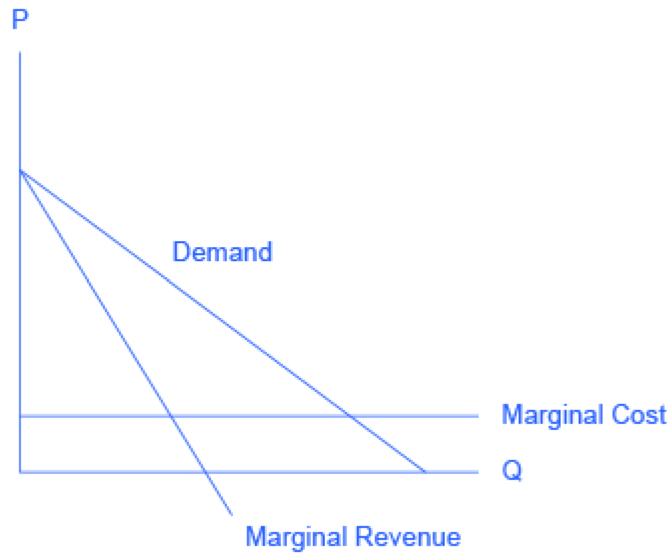
than self-interested behavior. However, the challenge for the parties is to find ways to encourage cooperative behavior.

Self-Check Questions

Exercise:

Problem:

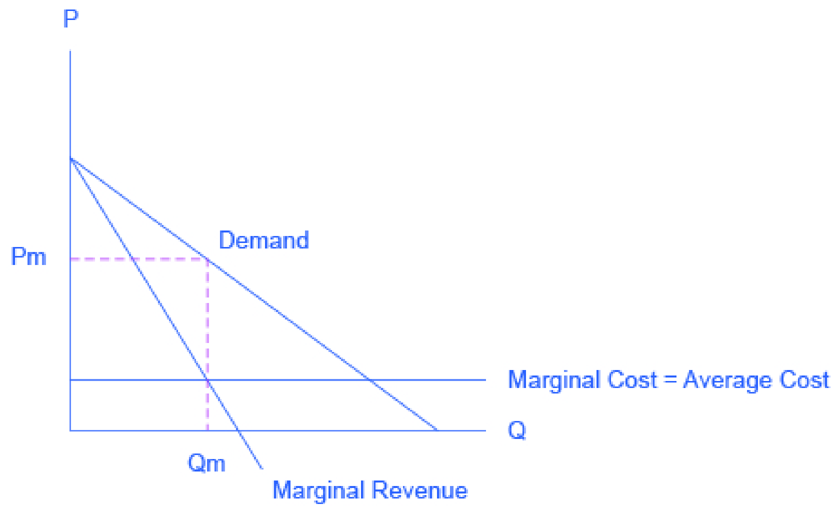
Consider the curve shown in [link], which shows the market demand, marginal cost, and marginal revenue curve for firms in an oligopolistic industry. In this example, we assume firms have zero fixed costs.



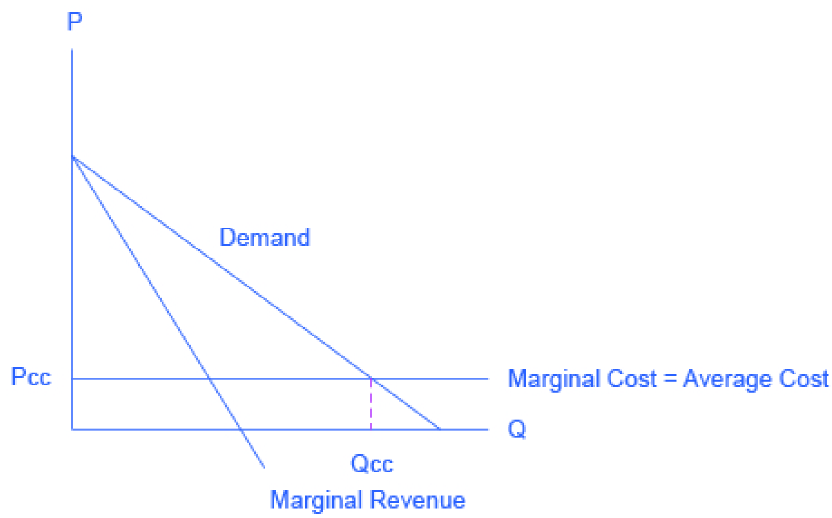
a. Suppose the firms collude to form a cartel. What price will the cartel charge? What quantity will the cartel supply? How much profit will the cartel earn?
b. Suppose now that the cartel breaks up and the oligopolistic firms compete as vigorously as possible by cutting the price and increasing sales. What will the industry quantity and price be? What will the collective profits be of all firms in the industry?
c. Compare the equilibrium price, quantity, and profit for the cartel and "cutthroat competition" outcomes.

Solution:

a. If the firms form a cartel, they will act like a monopoly, choosing the quantity of output where MR = MC. Drawing a line from the monopoly quantity up to the demand curve shows the monopoly price. Assuming that fixed costs are zero, and with an understanding of cost and profit, we can infer that when the marginal cost curve is horizontal, average cost is the same as marginal cost. Thus, the cartel will earn positive economic profits equal to the area of the rectangle, with a base equal to the monopoly quantity and a height equal to the difference between price (on the demand above the monopoly quantity) and average cost, as shown in the following figure.



b. The firms will expand output and cut price as long as there are profits remaining. The long-run equilibrium will occur at the point where average cost equals demand. As a result, the oligopoly will earn zero economic profits due to “cutthroat competition,” as shown in the next figure.



c. Pc > Pcc. Qc < Qcc. Profit for the cartel is positive and large. Profit for cutthroat competition is zero.

Exercise 2
Sometimes oligopolies in the same industry are very different in size. Suppose we have a duopoly where one firm (Firm A) is large and the other firm (Firm B) is small, as shown in the prisoner's dilemma box in Table 3. Assuming that the payoffs are known to both firms, what is the likely outcome in this case?

`Table 3: Prisoners' dilemma`

	Firm B colludes with Firm A	Firm B cheats by selling more output
Firm A colludes with Firm B	A gets R1,000, B gets R100	A getsR800, B gets R200
Firm A cheats by selling more output	A gets R1,050, B gets R50	A gets R500, B gets R20

Solution
Firm B reasons that if it cheats and Firm A does not notice, it will double its money. Since Firm A's profits will decline substantially, however, it is likely that Firm A will notice and if so, Firm A will cheat also, with the result that Firm B will lose 90% of what it gained by cheating. Firm A will

reason that Firm B is unlikely to risk cheating. If neither firm cheats, Firm A earns R1000. If Firm A cheats, assuming Firm B does not cheat, A can boost its profits only a little, since Firm B is so small. If both firms cheat, then Firm A loses at least 50% of what it could have earned. The possibility of a small gain (R50) is probably not enough to induce Firm A to cheat, so in this case it is likely that both firms will collude.

Review Questions

Exercise 3
Will the firms in an oligopoly act more like a monopoly or more like competitors? Briefly explain.

Exercise 4
Does each individual in a prisoner’s dilemma benefit more from cooperation or from pursuing self-interest? Explain briefly.

Exercise 5
What stops oligopolists from acting together as a monopolist and earning the highest possible level of profits?

Critical Thinking Questions

Exercise 6
Would you expect the kinked demand curve to be more extreme (like a right angle) or less extreme (like a normal demand curve) if each firm in the cartel produces a near-identical product like OPEC and petroleum? What if each firm produces a somewhat different product? Explain your reasoning

Exercise 7
When OPEC raised the price of oil dramatically in the mid-1970s, experts said it was unlikely that the cartel could stay together over the long term—that the incentives for individual members to cheat would become too strong. More than forty years later, OPEC still exists. Why do you think OPEC has been able to beat the odds and continue to collude? Hint: You may wish to consider non-economic reasons.

References

Benjamin, C. and De Wet, P. 2013. The flaw that broke the construction cartel's back. A R4.7-billion construction cabal was brought down by one player refusing to pay the others their dues. Mail and Guardian, July 19, sec. Business. Available http://mg.co.za/article/2013-07-19-00-the-flaw-that-broke-the-construction-cartels-back (Accessed: 22 May 2016)

Inggs, M. 2009. Coo-Ee soft drink brand helps Shoreline win market share. Business Day Live, August 7, sec. Archive. Available http://www.bdlive.co.za/articles/2009/08/07/coo-ee-soft-drink-brand-helps-shoreline-win-market-share-... (Accessed: 22 May 2016)

Smith, A. 1776. An inquiry into the nature and causes of the wealth of nations. London: W. Strahan and T. Cadell. 2 volumes.

Glossary

cartel
: a group of firms that collude to produce the monopoly output and sell at the monopoly price

collusion
: when firms act together to reduce output and keep prices high

duopoly
: an oligopoly with only two firms

game theory
: a branch of mathematics often used by economists that analyzes situations in which players must make decisions and then receive payoffs based on what decisions the other players make

kinked demand curve
: a perceived demand curve that arises when competing oligopoly firms commit to match price cuts, but not price increases

prisoner's dilemma

a game in which the gains from cooperation are larger than the rewards from pursuing self-interest

The Use of Mathematics in Principles of Economics

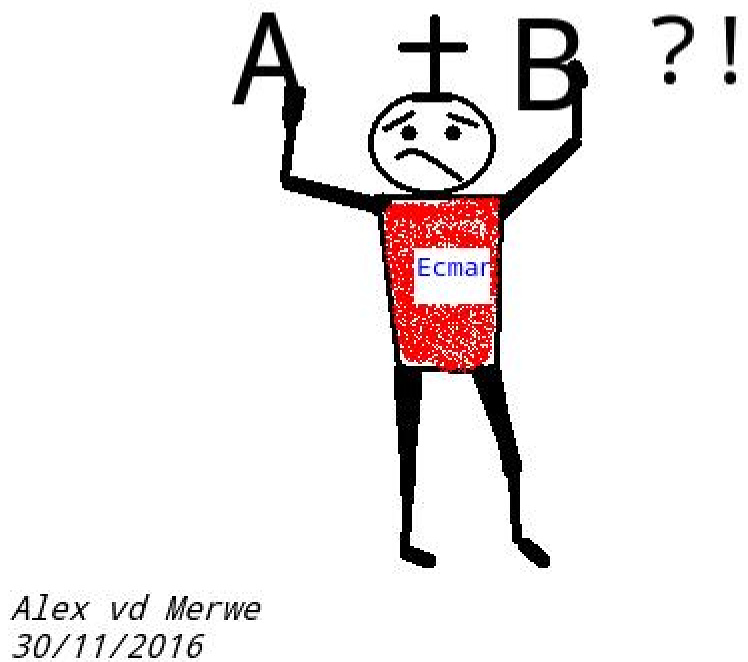


Figure 1. Maths scares many people but it is only a tool. (Credit: Alex van der Merwe)

Economics is not maths. There is no important concept in this course that cannot be explained without mathematics. However, maths is a tool that can be used to illustrate economic concepts. Remember the saying a picture is worth a thousand words? Instead of a picture, think of a graph. It is the same thing. Economists use models as the primary tool to explain and understand economic issues and problems. Maths is one way of working with (or manipulating) economic models.

There are other ways of representing models, such as text or explanations/stories. But why would you use your fist to bang a nail, if you had a hammer? Maths has certain advantages over text or verbal explanations/stories. It disciplines your thinking by making you specify exactly what you mean. You can get away with fuzzy thinking in your head, but you cannot when you use algebraic equations to produce models. At the same time, maths also has disadvantages. Mathematical models are necessarily based on simplifying assumptions, so they are not likely to be perfectly realistic. Mathematical models also lack some of the finer details which can be found in models that use writing/text, pictures or verbal explanations. The point is that maths is one tool, but it is not the only tool or even always the best tool economists can use. So what maths will you need for this book? The answer is: little more than high school algebra and graphs. You will need to know:

- What a function is
- How to interpret the equation of a line (i.e., slope and intercept)
- How to manipulate a line (i.e., changing the slope or the intercept)
- How to compute and interpret a growth rate (i.e., percentage change)
- How to read and manipulate a graph

In this text, we will use the easiest maths possible, and we will introduce it in this appendix. So if you find some maths in the book that you cannot follow, come back to this appendix to review. A little maths ability goes a long way; you **REALLY** do not need to know a great deal of maths to succeed in Economics. If you are going to major in economics, you should consider learning a little calculus. It will be worth your while in terms of helping you learn advanced economics more quickly.

Algebraic Models

Often economic models (or parts of models) are expressed in terms of mathematical functions. What is a function? "Function" refers to how something functions or works. A function describes a relationship. Sometimes the relationship is a definition. For example (using words), your lecturer is Mr Kabange. This could be expressed as Lecturer = Mr Kabange. Or Friends = Siwe + Nombuso + Eric.

Often in economics, functions describe cause and effect. The variable on the left-hand side is what is being explained (“the effect”). A "variable" is a word or symbol that can take on any value. On the right-hand side is what is doing the explaining (“the causes”). For example, you could explain the way your ENCS111 Final Mark is calculated using an equation like this:

ENCS111 Final Mark = (0.4 × Course Mark) + (0.6 × Exam Mark)

This equation states that your ENCS111 Final Mark depends on two things: your Course Mark (but only 40% of it) plus your Exam Mark (but only 60% of it). If this relationship is true, how could you raise your ENCS111 Final Mark? By not skipping classes, studying hard and scoring as many marks as you can not only in the final examination but also in the class tests that make up the Course Mark.

Economic models express relationships using economic variables, like Budget = money_spent_on_econ_books + money_spent_on_music, assuming that the only things you buy are economics books and music.

Most of the relationships we use in this course are expressed as linear equations such as this one:
Equation:

$$y = b + mx$$

Expressing Equations Graphically

Graphs are useful for two purposes. The first is to express equations visually, and the second is to display statistics or data. This section will discuss expressing equations visually.

To a mathematician or an economist, a variable is the name given to a quantity that may assume a range of values. In the linear equation presented above (y=b+mx), x and y are the variables, with x on the horizontal axis and y on the vertical axis, and b and m represent factors that determine the shape of the line. To see how this equation works, consider a numerical example:

Equation:

$$y = 9 + 3x$$

In this equation for a specific line, the b term has been set equal to 9 and the m term has been set equal to 3. Table 1 shows the values of x and y for this given equation. Figure 2 shows this equation, and these values, in a graph. To construct the table, just plug in a series of different values for x, and then calculate what value of y results. In the figure, these points are plotted and a line is drawn through them.

x	**y**
0	9
1	12
2	15
3	18
4	21
5	24
6	27

Values for the Slope Intercept Equation

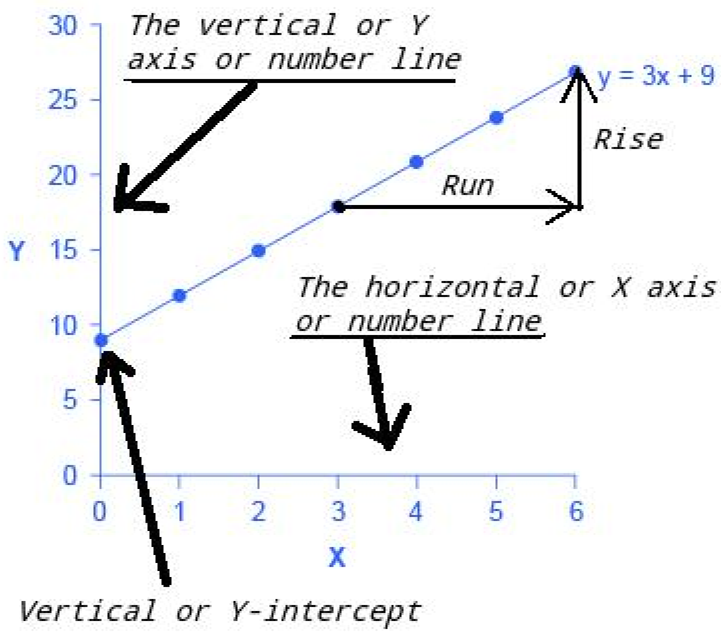


Figure 2: This line graph has x on the horizontal axis and y on the vertical axis. The y-intercept—that is, the point where the line intersects the y-axis—is 9. The slope of the line is 3; that is, there is a rise of 3 on the vertical axis for every increase of 1 on the horizontal axis. The slope is the same all along a straight line.

This example illustrates how the b and m terms in an equation for a straight line determine the shape of the line. The b term is called the y-intercept. The reason for this name is that, if x = 0, then the b term will reveal where the line intercepts, or crosses, the y-axis. In this example, the line hits the vertical axis at 9. The m term in the equation for the line is the slope. Remember that slope is defined as rise over run; more specifically, the slope of a line from one point to another is the change in the vertical axis divided by the change in the horizontal axis. In this example, each time the x term increases by one (the run), the y term rises by three. Thus, the slope of this line is three. Specifying a y-intercept and a slope—that is, specifying b and m in the equation for a line—will identify a specific line. Although it is rare

for real-world data points to arrange themselves as an exact straight line, it often turns out that a straight line can offer a reasonable approximation of actual data.

Interpreting the Slope

The concept of slope is very useful in economics, because it measures the relationship between two variables. A positive slope means that two variables are positively related; that is, when x increases, so does y, or when x decreases, y decreases also. Graphically, a positive slope means that as a line on the line graph moves from left to right, the line rises. The length-weight relationship, shown in Figure 5 later in this Appendix, has a positive slope. We will learn in other chapters that price and quantity supplied have a positive relationship; that is, firms will supply more when the price is higher.

A negative slope means that two variables are negatively related; that is, when x increases, y decreases, or when x decreases, y increases. Graphically, a negative slope means that, as the line on the line graph moves from left to right, the line falls. The altitude-air density relationship, shown in Figure 6 later in this appendix, has a negative slope. We will learn that price and quantity demanded have a negative relationship; that is, consumers will purchase less when the price is higher.

A slope of zero means that there is no relationship between x and y. Graphically, the line is flat; that is, zero rise over the run. Figure 7 of the unemployment rate, shown later in this appendix, illustrates a common pattern of many line graphs: some segments where the slope is positive, other segments where the slope is negative, and still other segments where the slope is close to zero.

The slope of a straight line between two points can be calculated in numerical terms. To calculate slope, begin by designating one point as the "starting point" and the other point as the "end point" and then calculating the rise over run between these two points. As an example, consider the slope of the air density graph (Figure 6 which is based on Table 3) between the points representing an altitude of 4,000 meters and an altitude of 6,000 meters:

Rise: Change in variable on vertical axis (end point minus original point)
Equation:

$$
\begin{aligned}
&= 0.100 - 0.307 \\
&= -0.207
\end{aligned}
$$

Run: Change in variable on horizontal axis (end point minus original point)
Equation:

$$
\begin{aligned}
&= 6{,}000 - 4{,}000 \\
&= 2{,}000
\end{aligned}
$$

Thus, the slope of a straight line between these two points would be that from the altitude of 4,000 meters up to 6,000 meters, the density of the air decreases by approximately 0.1 kilograms/cubic meter for each of the next 1,000 meters

Suppose the slope of a line were to increase. Graphically, that means it would get steeper. Suppose the slope of a line were to decrease. Then it would get flatter. These conditions are true whether or not the slope was positive or negative to begin with. A higher positive slope means a steeper upward rise of the line, while a smaller positive slope means a flatter upward rise the line. A negative slope that is larger in absolute value (that is, more negative) means a steeper downward drop of the line. A slope of zero is a horizontal flat line. A vertical line has an infinite slope.

Slope indicates not only the **type** of relationship (positive/direct or negative/inverse) between two variables but also the **strength** of that relationship. The steeper the slope/gradient, the stronger the relationship and the flatter the slope/gradient, the weaker the relationship between the X and Y variables. So in Figure 3 on the steep graph, for instance, it takes just a small decrease in the X variable (negative **run**, shown by the red arrow)) to produce a **rise** from A to B in variable Y. However, on the flat graph it takes quite a big increase in the X variable (shown y the blue arrow) to produce the same rise in variable Y from A to B. The flat graph shows a

weak, positive (direct) relationship between X and Y. The steep graph, oth the other hand, reflects a strong, negative relationship between X and Y

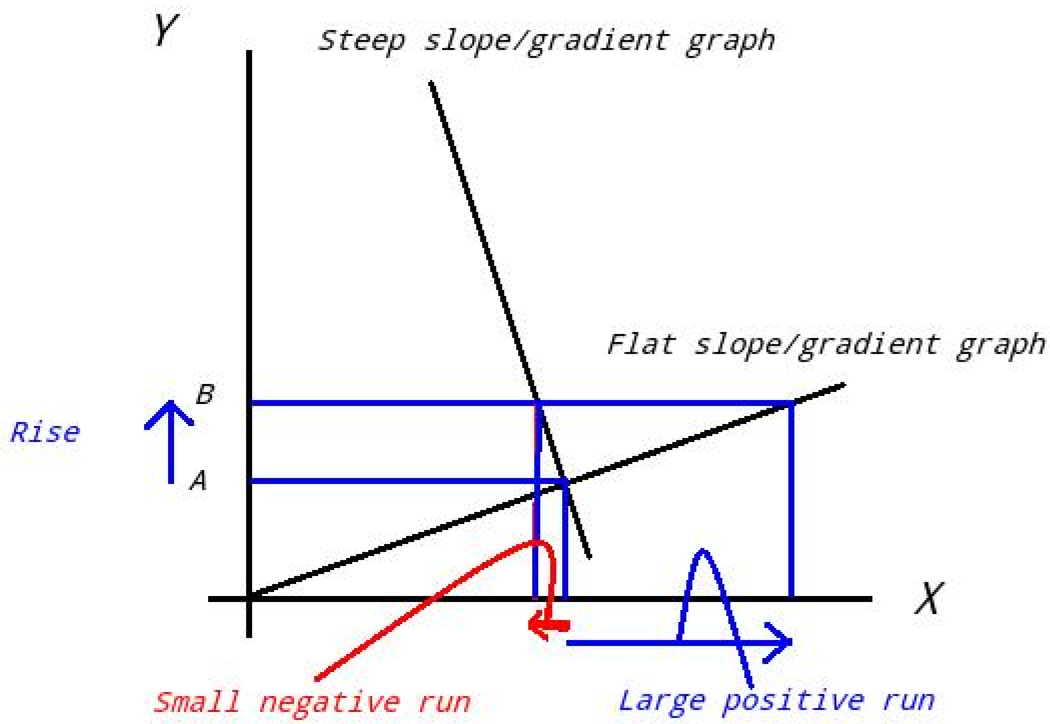


Figure 3. Slope indicates type and strength of relationship between X and Y.

Suppose a line has a larger intercept. Graphically, that means it would shift out (or up) from the old origin, parallel to the old line. If a line has a smaller intercept, it would shift in (or down), parallel to the old line.

Solving Models with Algebra

Economists often use models to answer a specific question, such as: What will the unemployment rate be if the economy grows at 3% per year? Answering specific questions requires solving the “system” of equations that represent the model.

Suppose the demand for cell phone sim cards is given by the following equation:
Equation:

$$\text{Qd} = 16 - 2\text{P}$$

where Qd (quantity demanded) is the amount of sim cards consumers want to buy, and P is the price of sim cards. Suppose the supply of sim cards is:
Equation:

$$\text{Qs} = 2 + 5\text{P}$$

where Qs (quantity supplied) is the amount of sim cards producers will supply.

Finally, suppose that the sim card market operates where supply equals demand, or
Equation:

$$\text{Qd} = \text{Qs}$$

We now have a system of three equations and three unknowns (Qd, Qs, and P), which we can solve with algebra:

Since at the market equilibrium price Qd = Qs, we can set the demand and supply equations equal to each other:
Equation:

$$\begin{aligned} \text{Qd} &= \text{Qs} \\ 16 - 2\text{P} &= 2 + 5\text{P} \end{aligned}$$

Subtracting 2 from both sides and adding 2P to both sides yields:
Equation:

$$\begin{aligned} 16 - 2P - 2 &= 2 + 5P - 2 \\ 14 - 2P &= 5P \\ 14 - 2P + 2P &= 5P + 2P \\ 14 &= 7P \\ \frac{14}{7} &= \frac{7P}{7} \\ 2 &= P \end{aligned}$$

In other words, the price of each sim card will be R2. How much will consumers buy?

Taking the price of R2, and inserting it into the demand equation, we get:
Equation:

$$\begin{aligned} Qd &= 16 - 2P \\ &= 16 - 2(2) \\ &= 16 - 4 \\ &= 12 \end{aligned}$$

So if the price is R2 per sim card, consumers will purchase 12 sim cards. How much will producers supply? Taking the price of R2, and inserting it into the supply equation, we get:
Equation:

$$\begin{aligned} Qs &= 2 + 5P \\ &= 2 + 5(2) \\ &= 2 + 10 \\ &= 12 \end{aligned}$$

So if the price is R2 each, producers will supply 12 sim cards. This means we did our maths correctly, since Qd = Qs.

Solving Models with Graphs

If algebra is not your strong point, you can get the same answer by using graphs. Take the equations for Qd and Qs and graph them on the same set of axes as shown in Figure 4 below. Since P is on the vertical axis, it is easiest if you solve each equation for P. The demand curve is then $P = 8 – 0.5Qd$ and the supply curve is $P = –0.4 + 0.2Qs$. Note that the vertical intercepts are 8 and –0.4, and the slopes are –0.5 for demand and 0.2 for supply. If you draw the graphs carefully, you will see that where they cross (Qs = Qd), the price is R2 and the quantity is 12 sim cards, just like the algebra predicted.

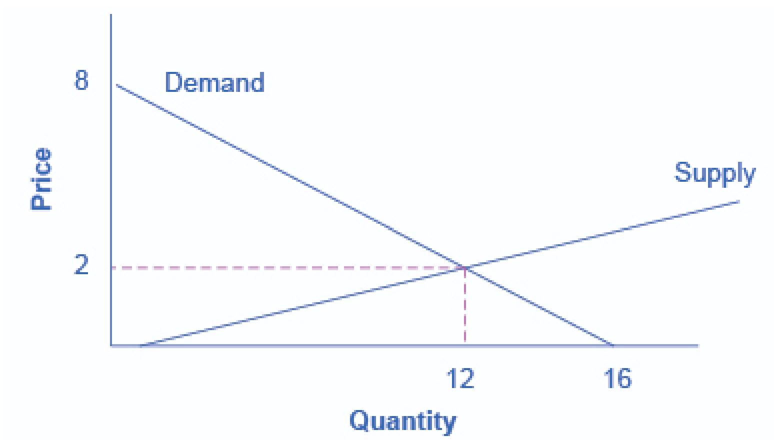


Figure 4: The equations for Qd and Qs are displayed graphically by the sloped lines.

We will use graphs more frequently in this book than algebra, but now you know the maths behind the graphs.

Growth Rates

Growth rates are frequently encountered in real world economics. A growth rate is simply the percentage change in some quantity. It could be your income. It could be a business's sales. It could be a nation's GDP. The formula for computing a growth rate is straightforward:

Equation:

$$\text{Percentage change} = \frac{\text{Change in quantity}}{\text{Quantity}}$$

Suppose your job pays R150 per hour. Your boss, however, is so impressed with your work that he gives you a R30 per hour raise. The percentage change (or growth rate) in your pay is R30/R150 = 0.20 or 20%.

To compute the growth rate for data over an extended period of time, for example, the average annual growth in GDP over a decade or more, the denominator is commonly defined a little differently. In the previous example, we defined the quantity as the initial quantity—or the quantity when we started. This is fine for a one-time calculation, but when we compute the growth over and over, it makes more sense to define the quantity as the average quantity over the period in question, which is defined as the quantity halfway between the initial quantity and the next quantity. This is harder to explain in words than to show with an example. Suppose South Africa's GDP was R1 trillion in 2005 and R1.03 trillion in 2006. The growth rate between 2005 and 2006 would be the change in GDP (R1.03 trillion – R1.00 trillion) divided by the average GDP between 2005 and 2006 (R1.03 trillion + R1.00 trillion)/2. In other words:

Equation:

$$\begin{aligned}
&= \frac{\text{R1.03 trillion} - \text{R1.00 trillion}}{(\text{R1.03 trillion} + \text{R1.00 trillion})\ /\ 2} \\
&= \frac{0.03}{1.015} \\
&= 0.0296 \\
&= 2.96\%\ \text{growth}
\end{aligned}$$

Note that if we used the first method, the calculation would be (R1.03 trillion – R1.00 trillion) / R1.00 trillion = 3% growth, which is

approximately the same as the second, more complicated method. If you need a rough approximation, use the first method. If you need accuracy, use the second method. Both methods of calculating percent change are used to compute elasticity coefficients in this textbook.

A few things to remember: A positive growth rate means the quantity is growing. A smaller growth rate means the quantity is growing more slowly. A larger growth rate means the quantity is growing more quickly. A negative growth rate means the quantity is decreasing.

The same increase over time yields a smaller growth rate. For example, if you got a R12,000 raise each year, in the first year the growth rate would be R12,000/R100,000 = 12%, as shown above. But in the second year, the growth rate would be R12,000/R112,000 = 10.71%. In the third year, the same R12,000 raise would correspond to a R12,000/R124,000 = 9.7%. The moral of the story is this: To keep the growth rate the same, the change must increase each period.

Displaying Data Graphically and Interpreting the Graph

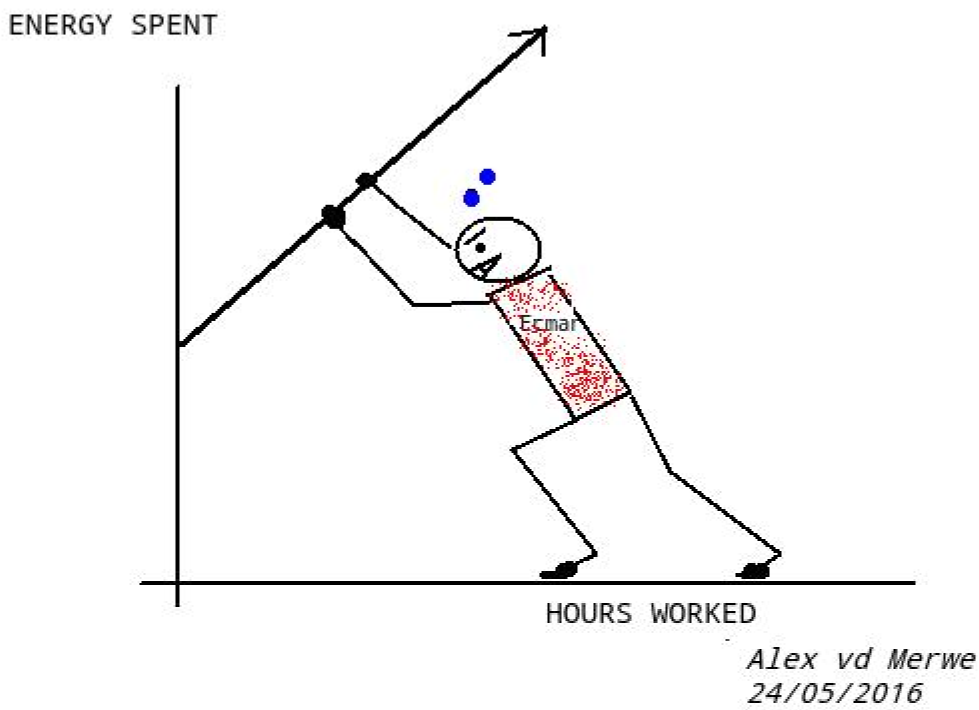


Figure 5. No need to fight with graphs Econman!

Graphs are also used to display data or evidence. Graphs are a method of presenting numerical patterns. They condense detailed numerical information into a visual form in which relationships and numerical patterns can be seen more easily. For example, which countries have larger or smaller populations? A careful reader could examine a long list of numbers representing the populations of many countries, but with over 200 nations in the world, searching through such a list would take concentration and time. Putting these same numbers on a graph can quickly reveal population patterns. Economists use graphs both for a simple and easily readable presentation of groups of numbers and for building a clear understanding of relationships and connections.

Three types of graphs are used in this book: line graphs, pie graphs, and bar graphs. Each is discussed below. We also provide warnings about how graphs can be manipulated to alter viewers' perceptions of the relationships in the data.

Line Graphs

The graphs we have discussed so far are called line graphs, because they show a relationship between two variables: one measured on the horizontal axis and the other measured on the vertical axis.

Sometimes it is useful to show more than one set of data on the same axes. We need to do this often in economics.The data in Table 2 is displayed in Figure 6 which shows the relationship between two variables: length and median weight for baby boys and girls during the first three years of life. (The median means that half of all babies weigh more than this and half weigh less.) The line graph measures length in centimetres on the horizontal axis and weight in kilograms on the vertical axis. For example, the reference lines on the graph show that a boy who is 66cm long will have a median weight of just over 7 kilograms (7.4 Kg). One line on the graph shows the length-weight relationship for boys and the other line shows the relationship for girls. This kind of graph is widely used by healthcare providers to check whether a child's physical development is roughly on track.

Length-weight relationship

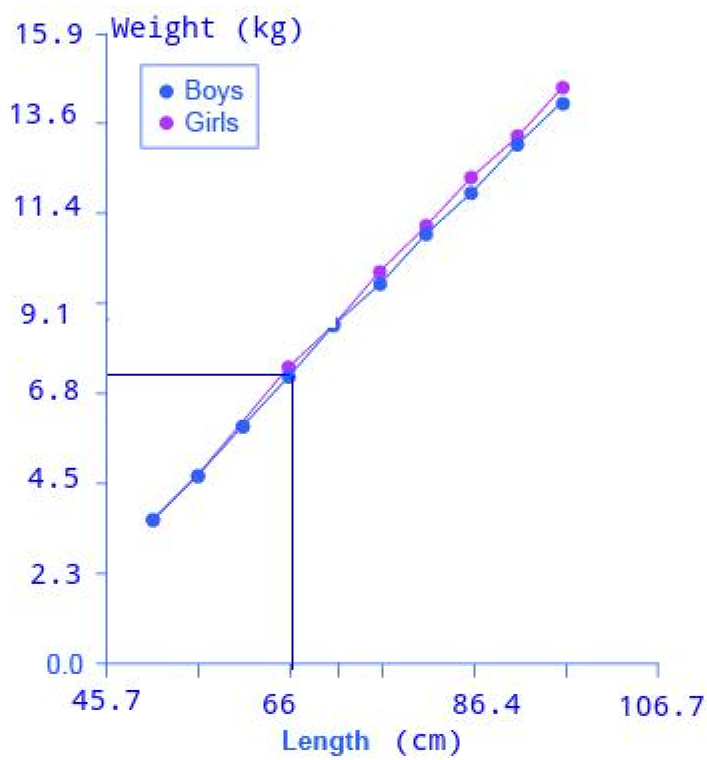


Figure 6: The line graph shows the relationship between height and weight for boys and girls from birth to 3 years. So, for instance, a boy of 66cm in height (measured on the horizontal axis) is typically just over 7kg (7.4kg) in weight (measured on the vertical axis). These data apply only to children in the first three years of life.

Boys from Birth to 36 Months		**Girls from Birth to 36 Months**	
Length (cm)	Weight (kg)	Length (cm)	Weight (kg)
50.8	3.6	50.8	3.6
55.9	4.8	55.9	4.8

Boys from Birth to 36 Months		Girls from Birth to 36 Months	
61	6.1	61	6
66	7.4	66	7.3
71.1	8.6	71.1	8.5
76.2	9.9	76.2	9.6
81.3	11	81.3	10.9
86.4	12.2	86.4	11.9
91.4	13.3	91.4	13.1
96.5	14.5	96.5	14.2

Length to Weight Relationship for Boys and Girls

Not all relationships in economics are linear (or constant in their rates of change). Sometimes they are curves which means that the rate of change between variables may increase or decrease at different rates. Figure 7 presents another example of a line graph, representing the data from Table 3. In this case, the line graph shows how thin the air becomes when you climb a mountain. The horizontal axis of the figure shows altitude, measured in meters above sea level. The vertical axis measures the density of the air at each altitude. Air density is measured by the weight of the air in a cubic meter of space (that is, a box measuring one meter in height, width, and depth). As the graph shows, air pressure is heaviest at ground level and becomes lighter as you climb. Table 3 shows that a cubic meter of air at an altitude of 500 meters weighs approximately one kilogram (about 2.2 pounds). However, as the altitude increases, air density decreases. A cubic meter of air at the top of Mount Everest, at about 8,828 meters, would weigh only 0.023 kilograms. The thin air at high altitudes explains why breathing becomes difficult and many mountain climbers need to use oxygen tanks as they climb to the top of high mountains.

Altitude-air density relationship

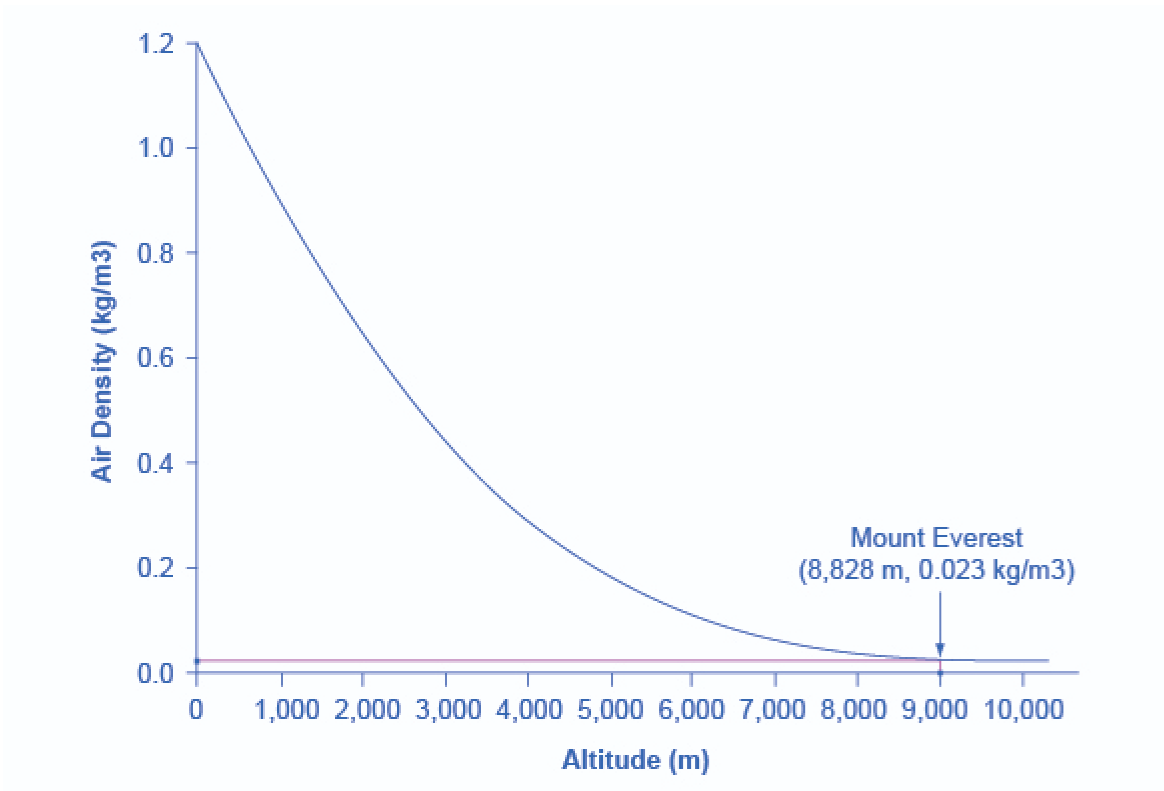


Figure 7: This line graph shows the relationship between altitude, measured in meters above sea level, and air density, measured in kilograms of air per cubic meter. As altitude rises, air density declines. The point at the top of Mount Everest has an altitude of approximately 8,828 meters above sea level (the horizontal axis) and air density of 0.023 kilograms per cubic meter (the vertical axis).

Altitude (meters)	Air Density (kg/cubic meters)
0	1.200

Altitude (meters)	**Air Density (kg/cubic meters)**
500	1.093
1,000	0.831
1,500	0.678
2,000	0.569
2,500	0.484
3,000	0.415
3,500	0.357
4,000	0.307
4,500	0.231
5,000	0.182
5,500	0.142
6,000	0.100
6,500	0.085
7,000	0.066
7,500	0.051
8,000	0.041
8,500	0.025

Altitude (meters)	Air Density (kg/cubic meters)
9,000	0.022
9,500	0.019
10,000	0.014

Altitude to Air Density Relationship

The length-weight relationship and the altitude-air density relationships in these two figures represent averages. If you were to collect actual data on air pressure at different altitudes, the same altitude in different geographic locations will have slightly different air density, depending on factors like how far you are from the equator, local weather conditions, and the humidity in the air. Similarly, in measuring the height and weight of children for the previous line graph, children of a particular height would have a range of different weights, some above average and some below. In the real world, this sort of variation in data is common. The task of a researcher is to organize that data in a way that helps to understand typical patterns. The study of statistics, especially when combined with computer statistics and spreadsheet programs, is a great help in organizing this kind of data, plotting line graphs, and looking for typical underlying relationships. For most economics and social science majors, a statistics course will be required at some point.

One common line graph is called a time series, in which the horizontal axis shows time and the vertical axis displays another variable. Thus, a time series graph shows how a variable changes over time. The graph below (Figure 8) shows the official unemployment rate in South Africa between 1999 and 2013. Unemployment is defined as the percentage of adults who want jobs and are looking for a job, but cannot find one. The points for the unemployment rate in each year are plotted on the graph, and a line then connects the points, showing how the unemployment rate has moved up and down since 1999. The line graph makes it easy to see, for example, that the highest unemployment rate during this time period was about 37% between 2001 and 2003.The unemployment rate decreased considerably between

2003 and 2004 but since then had declined only very slightly or remained constant. From 2012 to 2013 it seems that the rate of unemployment started to increase again.

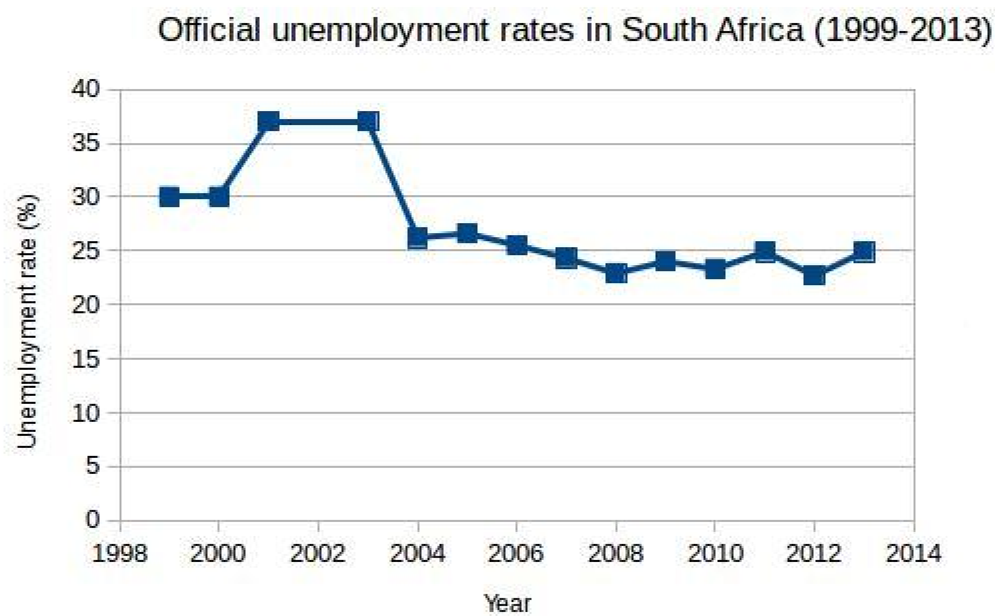


Figure 8. The line graph makes it easier to identify trends and patterns.

Pie Graphs

A pie graph (sometimes called a pie chart) is used to show how an overall total is divided into parts. A circle represents a group as a whole. The slices of this circular "pie" show the relative sizes of subgroups.

Figure 9 shows how the South Africa's estimated population was divided among children, working age adults, and the elderly in 2015. The information is first conveyed with numbers in Table 4, and then in the pie chart.

Year	Total Population (millions)	Under 15 (%)	15–64 years (%)	At least 64 (%)
2005	47 793 000	31.7	64.2	4.1
2010	50 133 000	30.1	65.2	4.6
2015	54 490 000	29.2	63	7.7

Table 4. South Africa's population/age distribution in 2005, 2010 and estimated in 2015 (Source: Wikipedia, 2016)

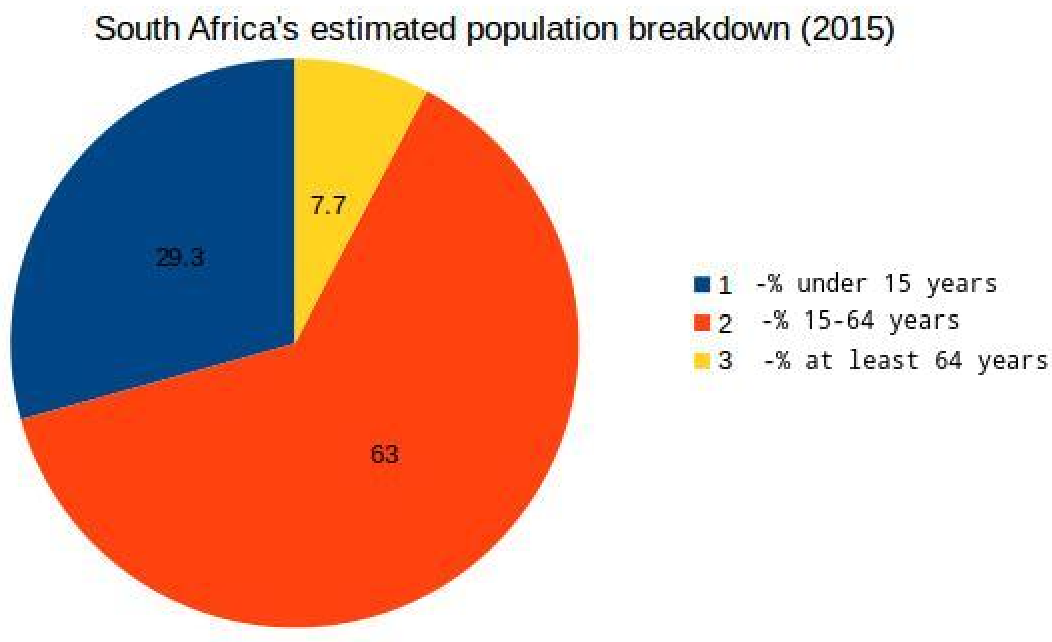


Figure 9. The pie graph illustrates South Africa's age distribution estimated for 2015.

In a pie graph, each slice of the pie represents a share of the total, or a percentage. For example, 50% would be half of the pie and 20% would be one-fifth of the pie. The pie graph allows you to get a feel for the relative size of the different age groups in 2015 without requiring you to slog through the specific numbers and percentages in the table. Some common examples of how pie graphs are used include dividing the population into groups by age, income level, ethnicity, religion, occupation; dividing different firms into categories by size, industry, number of employees; and dividing up government spending or taxes into its main categories.

Bar Graphs

A bar graph uses the height of different bars to compare quantities. Table 5 lists the 12 most populous countries in the world. Figure 10 provides this same data in a bar graph. The height of the bars corresponds to the population of each country. Although you may know that China and India are the most populous countries in the world, seeing how the bars on the graph tower over the other countries helps illustrate the magnitude of the difference between the sizes of national populations.

Leading Countries of the World by Population, 2015 (in millions)

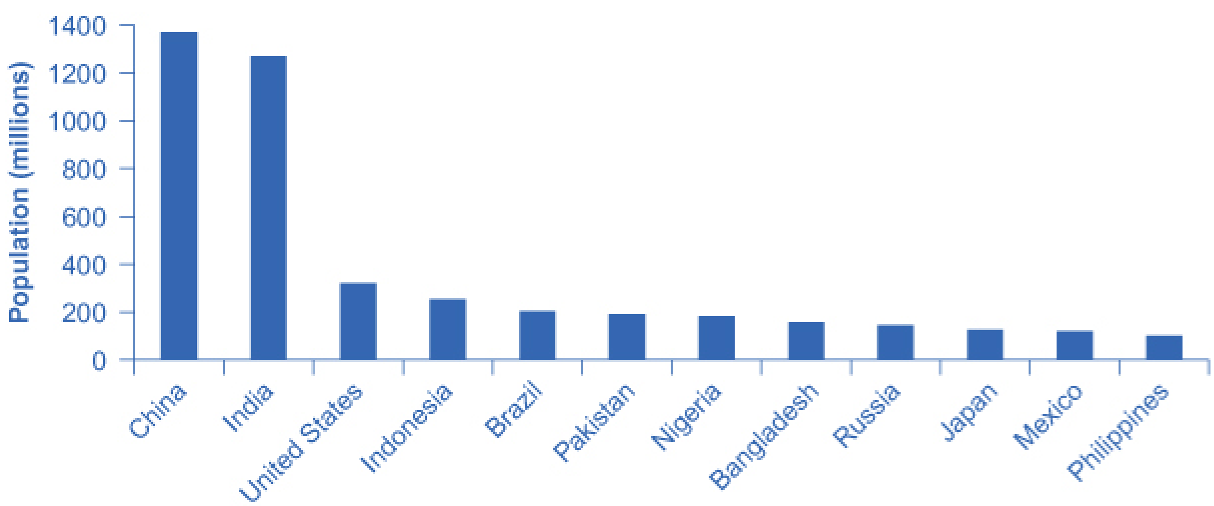


Figure 10. The graph shows the 12 countries of the world

with the largest populations. The height of the bars in the bar graph shows the size of the population for each country.

Country	Population (millions)
China	1369
India	1270
United States	321
Indonesia	255
Brazil	204
Pakistan	190
Nigeria	184
Bangladesh	158
Russia	146
Japan	127
Mexico	121
Philippines	101

Table 5: Leading 12 Countries of the World by Population

Comparing Line Graphs with Pie Charts and Bar Graphs

Now that you are familiar with pie graphs, bar graphs, and line graphs, how do you know which graph to use for your data? Pie graphs are often better than line graphs at showing how an overall group is divided. However, if a pie graph has too many slices, it can become difficult to interpret.

Bar graphs are especially useful when comparing quantities. For example, if you are studying the populations of different countries, as in Table 5, bar graphs can show the relationships between the population sizes of multiple countries. Not only can it show these relationships, but it can also - with some special adjustments - show breakdowns of different groups within the population.

A line graph is often the most effective format for illustrating a relationship between two variables that are both changing. For example, time series graphs can show patterns as time changes, like the unemployment rate over time. Line graphs are widely used in economics to present continuous data about prices, wages, quantities bought and sold, the size of the economy.

Being able to read graphs is an essential skill, both in economics and in life. A graph is just one perspective or point of view, shaped by choices such as those discussed in this section. Do not always believe the first quick impression from a graph. View with caution.

In economics we are often examining relationships between variables and looking for patterns of behaviour. Sometimes we need to report how one variable is affected by another. We may need to measure the size of the change in one variable caused by another. Sometimes the effect of one variable on some outcome may be greater or smaller than the effects produced by other variables. It would be useful to be able to report these patterns and relationships quickly and conveniently. Mathematical symbols help us to do this. We mostly use only three 3 or 4 mathematical symbols in this course. Econman below is learning about them in his own lesson.

Some funny symbols

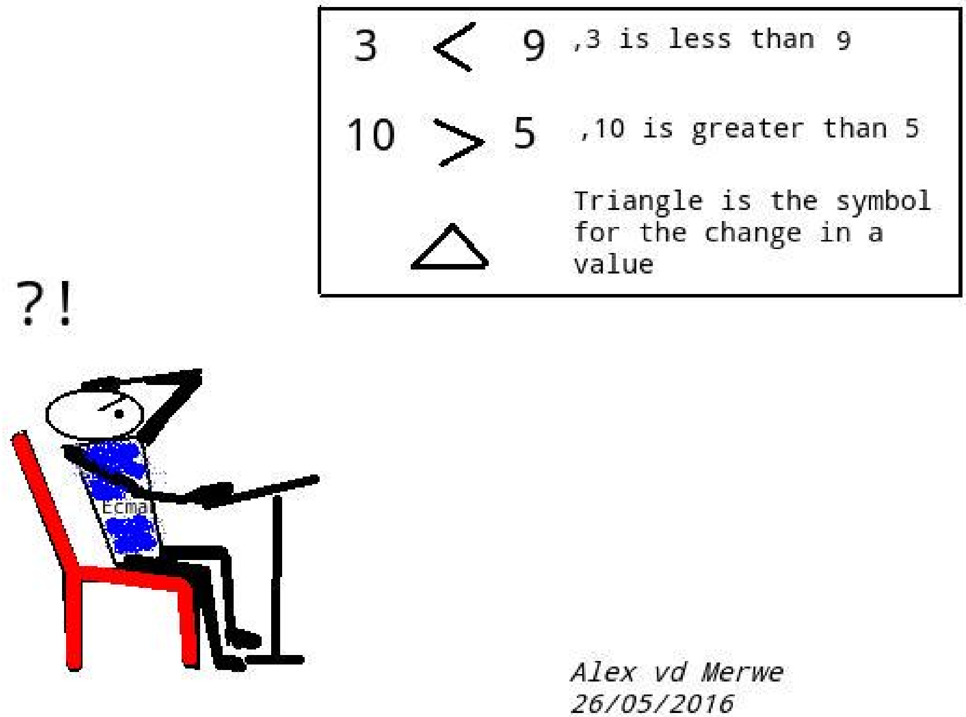


Figure 11. Don't panic Econman!!

Key Concepts and Summary

Maths is a tool for understanding economics. Economic relationships can be expressed mathematically using algebra or graphs. The algebraic equation for a line is y = b + mx, where x is the variable on the horizontal axis and y is the variable on the vertical axis, the b term is the y-intercept and the m term is the slope. The slope of a line is the same at any point on the line and it indicates the relationship (positive, negative, or zero) between two economic variables.

Economic models can be solved algebraically or graphically. Graphs allow you to illustrate data visually. They can illustrate patterns, comparisons, trends, and apportionment by condensing the numerical data and providing a clearer sense of relationships in the data. A line graph shows the relationship between two variables: one is shown on the horizontal axis and

one on the vertical axis. A pie graph shows how something is allotted, such as a sum of money or a group of people. The size of each slice of the pie is drawn to represent the corresponding percentage of the whole. A bar graph uses the height of bars to show a relationship, where each bar represents a certain entity, like a country or a group of people. The bars on a bar graph can also be divided into segments to show subgroups.

Any graph offers a picture of some relationship. The impression it leaves will be based on many choices, such as what data or time frame is included, how data or groups are divided up, the relative size of vertical and horizontal axes, whether the scale used on a vertical starts at zero. Thus, any graph should be regarded somewhat suspiciously, remembering that the underlying relationship can be open to different interpretations.

Review Questions

Exercise:

Problem:

Name three kinds of graphs and briefly state when is most appropriate to use each type of graph.

Exercise:

Problem: What is slope on a line graph?

Exercise:

Problem: What do the slices of a pie chart represent?

Exercise:

Problem: Why is a bar chart the best way to illustrate comparisons?

Exercise:

Problem:

How does the appearance of positive slope differ from negative slope and from zero slope?

References

Wikipedia. 2016. “Demographics of South Africa.” Available https://en.wikipedia.org/wiki/Demographics_of_South_Africa#Age_and_population_estimates:_1950_to_2010. (Accessed 30 Nov, 2016)